

**REPORT DOCUMENTATION PAGE**

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE July 2000	3. REPORT TYPE AND DATES COVERED 26-27 July 2001 Conference Proceedings - Final Report	
4. TITLE AND SUBTITLE  Input/Output and Imaging Technologies II Held in Taipei, Taiwan on 26-27 July 2000.			5. FUNDING NUMBERS	
6. AUTHOR(S)  Yung-Sheng Liu and Thomas s. Huang, Editors				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Industrial Technology Research Institute Taipei, Taiwan			8. PERFORMING ORGANIZATION REPORT NUMBER  ISSN 0277-786X	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) US Department of the Air Force Asian Office of Aerospace Research and Development (AOARD) Unit 45002 APO AP 96337-5002			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES Proceedings of SPIE, Vol. 4080. Published by: SPIE-The International Society for Optical Engineering, P.O. Box 10, Bellingham, Washington 98227-0010. This work relates to Department of the Air Force grant issued by the Asian Aerospace Office of Research and Development. The United States has a royalty free license throughout the world in all copyrightable material contained herein.				
12a. DISTRIBUTION/AVAILABILITY STATEMENT  Approved for Public Release. U.S. Government Rights License. All other rights reserved by the copyright holder. (Code 1, 2 <del>0</del> )			12b. DISTRIBUTION CODE  A	
12. ABSTRACT (Maximum 200 words)  This second Conference on Input/Output and Imaging Technologies, part of the International Optoelectronics Symposium Held in Taipei, Taiwan on 26-28 July 200 in conjunction with Photonics Taiwan 2000 (See <a href="http://www.spie.org/web/meetings/programs/pt00/pt00_home.html">http://www.spie.org/web/meetings/programs/pt00/pt00_home.html</a> ). This year's conference themes are: 3D object representation and vision-based applications, output devices and imaging, digital camera design and applications, and color imaging.				
13. SUBJECT TERMS AOARD, Foreign reports, 3D imaging, Digital cameras			15. NUMBER OF PAGES	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT  UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE  UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT  UNCLASSIFIED	20. LIMITATION OF ABSTRACT  UL	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)  
Prescribed by ANSI Std. Z39-18  
298-102



PROCEEDINGS OF SPIE  
SPIE—The International Society for Optical Engineering

# ***Input/Output and Imaging Technologies II***

**Yung-Sheng Liu  
Thomas S. Huang**  
*Chairs/Editors*

**26-27 July 2000  
Taipei, Taiwan**

*Sponsored by*  
SPIE—The International Society for Optical Engineering  
National Science Council (Taiwan)  
PIDA—Photonics Industry Development Association

*Published by*  
SPIE—The International Society for Optical Engineering



**20011130 080**

**Volume 4080**

SPIE is an international technical society dedicated to advancing engineering and scientific applications of optical, photonic, imaging, electronic, and optoelectronic technologies.

*AQ F02-02-0300*



The papers appearing in this book compose the proceedings of the technical conference cited on the cover and title page of this volume. They reflect the authors' opinions and are published as presented, in the interests of timely dissemination. Their inclusion in this publication does not necessarily constitute endorsement by the editors or by SPIE. Papers were selected by the conference program committee to be presented in oral or poster format, and were subject to review by volume editors or program committees.

Please use the following format to cite material from this book:

Author(s), "Title of paper," in *Input/Output and Imaging Technologies II*, Yung-Sheng Liu, Thomas S. Huang, Editors, Proceedings of SPIE Vol. 4080, page numbers (2000).

ISSN 0277-786X  
ISBN 0-8194-3719-0

Published by  
**SPIE—The International Society for Optical Engineering**  
P.O. Box 10, Bellingham, Washington 98227-0010 USA  
Telephone 1 360/676-3290 (Pacific Time) • Fax 1 360/647-1445  
<http://www.spie.org/>

Copyright ©2000, The Society of Photo-Optical Instrumentation Engineers.

Copying of material in this book for internal or personal use, or for the internal or personal use of specific clients, beyond the fair use provisions granted by the U.S. Copyright Law is authorized by SPIE subject to payment of copying fees. The Transactional Reporting Service base fee for this volume is \$15.00 per article (or portion thereof), which should be paid directly to the Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923 USA. Payment may also be made electronically through CCC Online at <http://www.directory.net/copyright/>. Other copying for republication, resale, advertising or promotion, or any form of systematic or multiple reproduction of any material in this book is prohibited except with permission in writing from the publisher. The CCC fee code is 0277-786X/00/\$15.00.

Printed in the United States of America.

# Contents

- vii *Conference Committee*
- ix *Introduction*

---

## KEYNOTE PAPER

---

- 2 **Photonic technologies in the 21st century: creation of new industries** [4080-201]  
T. Hiruma, Hamamatsu Photonics K.K. (Japan)

---

## SESSION 1 3D OBJECT REPRESENTATION AND VISION-BASED APPLICATIONS

---

- 8 **Progressive representation, transmission, and visualization of 3D objects (Invited Paper)** [4080-01]  
M. Okuda, T. Chen, Carnegie Mellon Univ. (USA)
- 14 **3D surface digitizing and modeling development at ITRI (Invited Paper)** [4080-02]  
W.-J. Hsueh, Industrial Technology Research Institute (Taiwan)
- 21 **Surface reconstruction technique based on 3D triangulation enhancement** [4080-03]  
I.-C. Chang, B.-T. Chen, K.-J. Hsieh, W.-J. Hsueh, H.-C. Lin, Industrial Technology Research Institute (Taiwan)
- 29 **Facial model estimation (FME) algorithms using stereo/mono image sequence** [4080-05]  
T.-G. Lin, Industrial Technology Research Institute (Taiwan); C. J. Kuo, National Chung Cheng Univ. (Taiwan)
- 41 **Vision-based intelligent robots** [4080-06]  
M.-C. Nguyen, Federal Armed Forces Univ./Munich (Germany)
- 48 **Face recognition security entrance** [4080-07]  
C. W. Ni, Industrial Technology Research Institute (Taiwan)
- 55 **Seagle-1: a new man-portable thermal imager** [4080-08]  
R.-N. Yeh, F.-F. Lu, H.-M. Hong, Y.-T. Cherng, H. Chang, Chung Shan Institute of Science and Technology (Taiwan)

---

## SESSION 2 OUTPUT DEVICES AND IMAGING

---

- 64 **New method of large ink supply without long tubing system for wide-format ink-jet printer** [4080-10]  
C.-T. Chen, Industrial Technology Research Institute (Taiwan)
- 72 **Meniscus oscillation of ink flow dynamics in thermal ink-jet print head** [4080-11]  
C.-L. Chiu, C.-W. Wang, Y.-Y. Wu, Y.-L. Lan, Industrial Technology Research Institute (Taiwan)



- 78 **Measurement of contrast ratios for 3D display [4080-12]**  
K.-C. Huang, Industrial Technology Research Institute (Taiwan); C.-H. Tsai, Industrial Technology Research Institute (Taiwan) and National Taiwan Univ.; K. Lee, W.-J. Hsueh, Industrial Technology Research Institute (Taiwan)
- 87 **Spatial long-range modulation of contrast discrimination [4080-13]**  
C.-C. Chen, Univ. of British Columbia (Canada); C. W. Tyler, Smith-Kettlewell Eye Research Institute (USA)

---

### SESSION 3 DIGITAL CAMERA DESIGN AND APPLICATIONS

---

- 96 **Measurement of the spatial frequency response (SFR) of digital still-picture cameras using a modified slanted-edge method [4080-14]**  
W.-F. Hsu, Y.-C. Hsu, K.-W. Chuang, Tatung Univ. (Taiwan)
- 104 **Comparisons of the camera OECF, the ISO speed, and the SFR of digital still-picture cameras [4080-15]**  
W.-F. Hsu, K.-W. Chuang, Y.-C. Hsu, Tatung Univ. (Taiwan)
- 112 **Digital camcorder image stabilizer based on gray-coded bit-plane block matching [4080-16]**  
Y.-M. Yeh, S.-J. Wang, H.-C. Chiang, National Chiao Tung Univ. (Taiwan) and Industrial Technology Research Institute (Taiwan)

---

### SESSION 4 COLOR IMAGING

---

- 122 **Internet color imaging (Invited Paper) [4080-17]**  
H.-C. Lee, Eastman Kodak Co. (USA)
- 136 **Spectral estimation and color appearance prediction of fluorescent materials [4080-18]**  
B.-K. Lee, F.-C. Shen, Chung Hua Univ. (Taiwan); C.-Y. Chen, Industrial Technology Research Institute (Taiwan)
- 148 **Design and production of color calibration targets for digital input devices [4080-19]**  
C. Wen, J. Lee, Industrial Technology Research Institute (Taiwan)
- 159 **Implementation of scanner ICC profile generator [4080-20]**  
Y.-C. Liaw, C.-Y. Chen, Industrial Technology Research Institute (Taiwan)
- 167 **Color reproduction system based on color appearance model and gamut mapping [4080-22]**  
F.-H. Cheng, C.-Y. Yang, Chung Hua Univ. (Taiwan)

---

### POSTER SESSION

---

- 180 **Characteristic extraction of face using DWT and recognition based on neural networks [4080-24]**  
H.-B. Kim, Chosun Univ. (Korea); C.-H. Lim, Dongkang College (Korea); S.-J. Park, Chunnam Univ. (Korea); J.-A. Park, Chosun Univ. (Korea)
- 192 **Some contributions to wavelet-based image coding [4080-25]**  
Y.-W. Li, K.-S. Chang, L.-S. Yan, D.-F. Shen, Yunlin Univ. of Science and Technology (Taiwan)

- 200 **Fast ITTBC using pattern code on subband segmentation [4080-26]**  
S. S. Koh, H. C. Kim, K. Y. Lee, H. B. Kim, Chosun Univ. (Korea); H. Jeong, G. S. Cho, Chosun College of Science and Technology (Korea); C. H. Kim, Chosun Univ. (Korea)
- 208 **Current research on ARO-positron emission tomography [4080-27]**  
M.-L. Jan, H.-C. Liang, S.-W. Huang, C.-S. Shyu, J.-S. Tang, H.-C. Liu, C.-C. Pei, C.-K. Yeh, Institute of Nuclear Energy Research (Taiwan)
- 214 **Influence of compression ratio of foam on printing quality of ink cartridge [4080-28]**  
C.-T. Chen, C.-C. Lai, Industrial Technology Research Institute (Taiwan)
- 222 **Influence of back pressure of ink cartridge on regular operation of ink supply system [4080-29]**  
C.-T. Chen, Industrial Technology Research Institute (Taiwan)
- 230 **Method and apparatus for measuring the droplet frequency response of an ink-jet print head [4080-30]**  
Z.-R. Lian, M.-L. Lee, Y.-H. Lai, H.-L. Hu, C. Wang, Industrial Technology Research Institute (Taiwan)
- 239 **Film stress and adhesion characteristics of passivation layers for thermal ink-jet print head [4080-31]**  
Y.-S. Lee, Y.-Y. Wu, C.-Y. Cheng, Industrial Technology Research Institute (Taiwan); D. S. Wu, Da-Yeh Univ. (Taiwan)
- 246 **Monolithic thermal ink-jet print head combining anisotropic etching and electroplating [4080-32]**  
C.-Y. Cheng, J.-P. Hu, Y.-H. Lai, H.-F. Wang, C.-T. Cheng, Industrial Technology Research Institute (Taiwan)
- 253 **Structured light based on shaped depth-image capturing system [4080-33]**  
D. Chen, G. Tan, Harbin Univ. of Science and Technology (China); X. Yu, Q. Meng, Harbin Engineering Univ. (China)
- 258 *Author Index*

## Conference Committee

### *Conference Chairs*

**Yung-Sheng Liu**, Industrial Technology Research Institute (Taiwan)  
**Thomas S. Huang**, University of Illinois/Urbana-Champaign (USA)

### *Program Committee*

**Chun-Yen Chen**, Industrial Technology Research Institute (Taiwan)  
**Masayoshi Esashi**, Tohoku University (Japan)  
**Shu-Cheng Hsieh**, Industrial Technology Research Institute (Taiwan)  
**Wen-Jean Hsueh**, Industrial Technology Research Institute (Taiwan)  
**Hiroshi Yasuda**, University of Tokyo (Japan)

### *Session Chairs*

- 1    3D Object Representation and Vision-Based Applications  
    **Wen-Jean Hsueh**, Industrial Technology Research Institute (Taiwan)
- 2    Output Devices and Imaging  
    **Shu-Cheng Shieh**, Industrial Technology Research Institute (Taiwan)
- 3    Digital Camera Design and Applications  
    **Wei-Feng Hsu**, Tatung University (Taiwan)
- 4    Color Imaging  
    **Chun-Yen Chen**, Industrial Technology Research Institute (Taiwan)

## **Introduction**

This second Conference on Input/Output and Imaging Technologies, part of the 2000 International Optoelectronics Symposium in Taiwan, marks the continuation of Taiwan's active participation in the input/output computer peripheral industry. It is our great pleasure to bring together researchers and developers in this area to spark discussions and interactions for further advancement.

Four sessions present the themes of this year's conference: 3D object representation and vision-based applications, output devices and imaging, digital camera design and applications, and color imaging. We have three invited speakers in the areas of 3D and color imaging, two fast-growing fields in image processing. The conference is comprised of 19 orally presented papers and 10 poster papers.

I would like to express my sincere appreciation to cochair Prof. Thomas S. Huang, to the program and organizing committees, and particularly to Dr. Wen-Jean Hsueh, for making this exciting conference possible. I believe that through active participation, all attendees will benefit significantly from this conference, and I urge all to take a more active role by submitting papers or participating in committees in the conferences to come.

**Yung-Sheng Liu**

## **Keynote Paper**

# Photonic Technologies in the 21st Century : Creation of New Industries \*

Teruo Hiruma  
Hamamatsu Photonics KK  
Hamamatsu, Japan

## ABSTRACT

As we approach the new millennium, the ongoing aim of human society is not only for promoting scientific technology but also creating new industries. To achieve this goal, each person in industry must recognize anew that the real meaning of science is to explore the absolute truth. It is also important that people recognize that there are unlimited matters which we humans do now yet know.

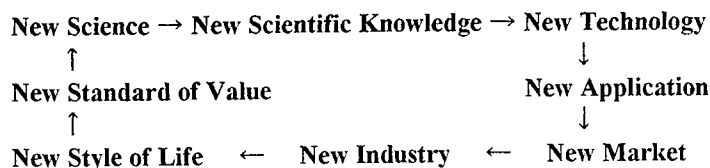
## 1. INTRODUCTION

The 20th century was one that witnessed many great discoveries and our knowledge increased many times during the last one hundred years. Yet even with such an explosion of knowledge and information there is much more that we do not yet understand. Our present knowledge represents only a fraction of what there is to know. For example, there have been major breakthroughs in understanding cell structure by studying individual components or systems such as the role of calcium ions in signal transduction. However very little is known about how all of these components work in concert. It is very much like trying to understand an orchestra by studying the individual instruments. In the future we will develop methods to study the function of the entire cell not just individual systems. On the molecular level we are just beginning to study the details of molecular dynamics during a chemical reaction. The work of this year's Nobel Prize winner in chemistry, Dr Zewail shows how it is possible to use photonics to study the intimate detail of a chemical reaction. Once we gain such detailed information about more complex systems it will be possible to more efficiently produce the chemicals we need and to destroy those that we no longer require.

At Hamamatsu Photonics it is our corporate mission to provide photonics technology that will help us to gain new knowledge of the world we live in. Photonic technologies are very unique in that they let us observe the parts of the world that are very far away (thousands of light years), very small (nanometers) or happen very quickly (in femtoseconds). The roots of our company can be found in the pioneering spirit of Professor Kenjiro Takayanagi who independently developed the technology of television despite the fact those around believed it could not be done because it had never been done. Professor Takayanagi hoped to develop a new way for people to experience the world. We inherited his spirit and continue this idea by using photonics to gain knowledge as well as improve the quality of life for all people.

In the twenty first century it will be possible that we could make all of mankind healthy. Not just in a physical sense but in the definition of the World Health Organization where "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." Photonics has the potential for creating the knowledge and industry that could make this possible. It is our hope that this new century sees the beginning of a new economic cycle shown in Figure 1. While industry is designed to generate profit, the purpose of industrialization is for all of mankind to share a common understanding of the New Life Style and to benefit from the New Standard of value, namely Health as defined above.

Figure 1:



\*Also published in Proceedings of SPIE Volumes 4078, 4079, 4081, and 4082

Through the application of photonics we are now beginning to develop the New Science that will lead to the cycle illustrated in Figure 1. In this cycle, mankind is constantly improving its status by using new technologies to discover new knowledge. The application of this new knowledge leads first to new industries and then to a change in the social fabric of society. We understand that this is a very long-term goal. But mankind needs to dream in order to progress. Only by trying to see over the horizon can we discover something that will radically improve all of our lives. Thus while Hamamatsu Photonics' short-term goal is to generate profits, these profits are to be used in the quest for new knowledge which is our long-term target.

## 2. TECHNOLOGIES

This paper will discuss several technologies that are key to the development of new knowledge which will eventually produce new industries. Application of these technologies will be also discussed.

### A. The Ultimate Laser Photon (Photon Factory)

The light emitted by a laser is unique in that it is monochromatic, coherent and directional. These properties have made laser-generated photons vital to all types of research ranging from biology to high-energy physics. We need to obtain a better understanding of exactly what is a photon and how it interacts with the world. By better understanding the photon on a fundamental level we will be able to use it more effectively. Phenomena such as the particle wave duality and teleportation must be better understood through a study of the photon.

The development of very small terawatt and petawatt laser systems give many researchers access to inexpensive ultra high power. New physical phenomena are being discovered when such intense laser beams interact with matter because these lasers create electric fields much greater than those seen in any other experiment do.

### B. Ultra Fast Measurement Technology

By continuing to push the speed at which we make measurements we will discover greater detail of how our world operates. We now have lasers that are capable of measuring the individual motion of atoms in molecules. Newer and faster methods will help get even greater detail of how molecules react. Even faster methods will allow us to follow the motion of electrons during important chemical reactions such as photosynthesis or vision.

### C. Optical Correlation Technology

Even though we have discovered only a small fraction of our knowledge, we are severely limited in using it because even this limited amount of knowledge is too great to process with conventional computer systems. We must learn how to process information in parallel with optical processors such as spatial light modulators. Ultimately our goal is to process information in 3 or more dimensions using technology that must still be invented.

### D. Forecast Simulators

With time not only has our information become too complex, but also the questions we need to answer become more difficult. As our planet's population increases and our technology becomes more complex, the risk of answering a question incorrectly grows exponentially. For example, the consequences of incorrectly predicting the outcome of global warming will be severe if we either under estimate or over estimate the significance of burning fossil fuels. Premature curtailing of fossil fuels will severely curtail the growth of developing countries leading to unnecessary pain and suffering. Failure to prevent global

warming will have even worse consequences. We need better methods to simulate events or conditions so we can better guide environmental, economic, technical, political, and military decisions.

Ultimately nations will never again fight a war on the battlefield but instead will use simulations to replace them. The simulations will simultaneously decide the output as well as convince the parties that physical conflict is too costly.

#### E. High Power Lasers

Photons are capable of doing many important things such as curing cancer, printing this manuscript or repairing an integrated circuit. At very high photon densities there are many new things that photons can do. At present, it is expensive to generate a lot of photons because the photon sources are expensive. Semiconductor laser diodes hold the promise of being able to reliably and inexpensively generate photons for many new and exciting applications. Just like the replacement of the vacuum tube with the transistor and then to the integrated circuit, so too will the semiconductor laser evolve and result in important technologies and new industries that we cannot even imagine today.

#### F. New Photochemistry

Much of our planet's energy is wasted in creating chemicals that we need to live or to improve our lives. Lasers are capable of creating specific excited states. Finding ways to selectively excite molecules so that they can be moved along specific reaction pathways will lead to huge savings in cost, energy and pollution. New knowledge on how to perform pathway specific photochemistry is vital to the goal of making everyone healthy according to the World Health Organization's definition of health.

### 3. APPLICATIONS

We can only speculate on what the full effect of such new photonic technologies will have in the long term. However over a short period of time we can easily imagine some of the benefits we might enjoy from these as well as other photonic technologies. Some of these benefits are discussed below.

#### A. Measurement of Physiological Functions

The pulse oximeter has already found an important role in guaranteeing that the oxygen concentration of the blood is maintained at as close to optimum as possible. Countless lives have been saved and others have had severe injury prevented by this simple optical device. Not very far away are devices that will permit rapid and painless screening for diabetes. Noninvasive cancer diagnosis is already being tested in clinical trials.

Ultimately a device will be available that checks your body's functions on a daily basis. It screens for potential problems before they cause disease. Adjustments to exercise, diet or even administration of drugs can be performed before the individual is aware of a problem. Such an advanced detection system would save costs, pain and anxiety. It would go as long way to attaining the goal of making people truly healthy.

#### B. Optical Medicine

In the past few years, photodynamic therapy has been shown to be a valuable treatment for some forms of cancer. In some cases it is far more useful than other techniques such as surgery because it leaves the effected organ in tact. Therefore for young women, cancer of the cervix no longer means that it is the end of their dream to have a family. For older people suffering from the wet form of macula degeneration, photodynamic therapy will soon be used to prevent the blindness caused by this disease.



New chemicals are being developed that are absorbed faster by the cancer cells and discharged more rapidly by the body. This will make treatment simpler and more effective. Patients may not even need to stay overnight in a hospital. Presently PDT can only be used on cancers that are found on a surface. Techniques are being developed that will be used in the treatment of cancers that are deep inside an organ.

Cosmetic uses of photons for hair removal, port wine stain removal or tattoo removal make it easier for a person to be accepted by society. These applications are far from superficial since they greatly improve the quality of life for those that need them.

Other applications of photonics to medical practice will certainly emerge in the near future for things such as the treatment of stroke, heart disease, healing of wounds and reducing or relieving pain.

### C. Early Detection of Disease

Cancer screening using Positron Emission Tomography (PET) holds the promise of early detection and cure of this terrible affliction. Injection of fluorodeoxyglucose into the blood stream is current used to uncover cells that are metabolizing at rates faster than those of their neighbors. These cells are then analyzed to determine if they are malignant. Such a screening method could in the near future make an entire city cancer death free.

Light CT uses nonionizing infrared photons to take a three dimensional image of the body. Work is under way in many places around the world to use light CT as a method for detecting breast cancer. This technique could be less expensive than x-ray methods and used safely on all individuals including pregnant women. Other uses of the light CT would be to quickly determine if a stroke is caused by ischemia or a hematoma. Such information is vital in determining the correct treatment. Rapid treatment of stroke can greatly reduce the damage to the brain resulting in a patient that can lead a normal life even after such a severe trauma.

In the future we hope to quantify the health of a person, not just the presence of disease.

### D. Fiberless Optical Communication

Information is the most important commodity in our society. We are constructing very large and expensive infrastructures to move information from one location to the other. Fiber optics is one of the key technologies for information transport because of the very high capacity available due to wavelength division multiplexing. This technique suffers from the fact that fibers must be placed between locations. At Hamamatsu Photonics we have developed a series of fiberless optical communications systems. These operate by transmitting the optical signals through air. They have the capability to send data, or video without the need for government licenses or owning a right of way. One such a fiberless system is used at sporting events such as golf tournaments to transmit the video camera output to the broadcaster's trailer or even back to a studio. Such a system was used at the Atlanta Olympics and is now being tested in Hamamatsu City. In our hometown it is being used to connect elementary schools with the city hall. It could also be used to connect remote clinics with the medical school for telemedicine.

### E. Health Industry for Successful Aging

Many countries will soon suffer from an increase in their average age. In the past such an increase in age would greatly burden society in terms of medical expenses and the cost of financially supporting an aging population. We believe that it is possible to completely eliminate the impact of a graying population by finding ways to reduce the pace and effect of the aging process. At Hamamatsu Photonics we are using photonic technology to understand how locomotion is effected as a person ages. We hope to develop exercises that will prevent the loss of mobility and greatly reduce the probability of an older person falling. While just a small step, it will have a big impact on the quality of life of our seniors.

#### F. Disposal of Industrial Waste

High power lasers and controlled photochemistry hold the promise of being able to safely dispose of dangerous waste products. It will do this by selective destruction of the dangerous ingredients into less danger or even harmless smaller molecules. These smaller molecules can then be recycled into new products.

#### G. Search for New Energy Sources

Perhaps the biggest impact that photonics can have on mankind is the development of clean and inexpensive energy. For once this is available; the quality of life of the entire world can be improved without damaging the planet. We must continue our search for a way to harness laser fusion and solar energy for they are needed to make the world a better place to live for all of us.

Photonics holds the promise of creating New Science and New Technology which will lead to New Industrial and of course to a world population that is truly healthy.

---

## **SESSION 1**

# **3D Object Representation and Vision-based Applications**

## Progressive Representation, Transmission, and Visualization of 3D Objects

Masahiro Okuda and Tsuhan Chen

Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Tel: +1 (412) 268-7536

Email: tsuhan@cmu.edu

Keywords: 3D data, 3D models, 3D meshes, geometry coding, texture coding, progressive coding, progressive transmission, streaming perceptual quality, visualization

### ABSTRACT

Files containing 3D objects, typically represented as 3D meshes with certain geometry and texture information, are very large. Therefore, not only do 3D objects take a lot of storage space, it is also extremely time-consuming to transmit them over the network for visualization. In addition, most 3D visualization applications need the entire 3D data file to render the 3D object even though the user may be interested in only a small part or a low-resolution version of the object. Progressive coding of 3D objects can resolve these problems. In this paper, we report our recent progress in progressive representation, transmission, and visualization of 3D objects. In our scheme, both geometry and the texture of the 3D object are progressively coded and transmitted. More perceptually important information is transmitted before the less important information, which allows the user to stop the transmission at any time and yet retain the best available perceptual quality of the object at that time. Furthermore, the visible portion of the object is transmitted first and the non-visible portion is transmitted later, or not transmitted at all, in order to save the overall bandwidth.

### 1. INTRODUCTION

Computer graphics using 3D objects are becoming more and more popular in many applications including movie productions, TV commercials, and video games. However, files containing 3D objects still remain to be very large, so it is time-consuming to retrieve 3D objects from the storage device or to download them from the network. Moreover, most 3D visualization applications have to obtain the entire file of a 3D model in order to display the model, even when the user is interested only in a small part, or a low-resolution version, of the model. This makes it very ineffective when 3D models need to be shared or transmitted over the network. Therefore, progressive representation of 3D objects is desired to solve these problems to meet various needs [1],[2],[3].

One existing scheme for 3D object representation is the compressed binary format specified by the Web3D Consortium, a group that standardizes formats of files containing 3D models [4]. Another example is the work recently done by the Synthetic/Natural Hybrid Coding (SNHC) subgroup with MPEG-4 [5],[6]. A 3D model generally contains geometry, texture, and other attribute data like normals and colors. Most of existing progressive coding algorithms focus only on the geometry. These algorithms are progressive either in terms of *resolution*, i.e., the number of vertices [7],[8], or in terms of the signal-to-noise ratio (SNR), i.e., the accuracy of vertex coordinates [9],[10],[11]. However, most existing schemes provide only a limited number of levels of detail (LOD). Furthermore, they consider only coding of the geometry information. Once some vertices and triangles are decimated during the simplification process, the corresponding attribute data are also discarded. To make the simplified model look realistic and similar to the original model, simplification of the attribute data including normals, colors and textures, should be taken into account together with the geometry.

In this paper, we propose a joint geometry/texture coding scheme for arbitrary manifold 3D models resulting in progressive bitstreams. The proposed algorithm is based on a vertex decimation approach. The 3D model is transmitted vertex-by-vertex,

providing “granular” progressive transmission. The textures are also coded progressively, and texture bits and geometry bits are combined into one bitstream. The proposed method allows the user to stop the transmission at any time and yet obtain the best available quality in terms of both geometry and texture. In addition to providing joint geometry/texture progressive coding, the proposed scheme is also comparable to or better than other existing schemes in terms of coding efficiency.

## 2. THE PROPOSED SCHEME

There are two ways to correspond attribute data with the 3D model. One is to associate attribute data to each vertex. In this paper, this type of data is called “vertex attributes” (vertex attributed texture coordinates, vertex attributed colors, and so on). The other to associate attribute data to each *corner*, i.e., each vertex in each triangle. We call this type of data “corner attributes” (corner attributed texture coordinates, corner attributed colors, and so on).

In the proposed scheme, the encoder removes vertices until we are left with a base mesh that has only a small fraction of the vertices and triangles in the original model. To send a 3D model progressively, we start by sending the vertex positions and vertex indices of the base mesh. Then, enhancements to this base mesh are sent, vertex-by-vertex, until all the vertices are transmitted. In the mean time, texture information is transmitted progressively, as detailed in the next section.

### 2.1. Vertex Decimation

The process of vertex decimation is as follows. In a triangular mesh, there is a ring of triangles that surround every vertex, as illustrated in Figure 2. To ensure that the most perceptually important vertices are sent before the less important vertices, we need a method for measuring the perceptual importance of vertices. We introduce two measures. The measure  $v(i)$  is defined as the difference in the volume caused by the decimation, by forming tetrahedrons with the removed vertex as the apex and the new triangles as the base, and adding up the volumes of these tetrahedrons. We can see that the measure  $v(i)$  is similar to the “curvature” of the mesh at the vertex.

The other measure  $c(i)$  is to quantify texture similarity in the surrounding triangles. During the simplification procedure, some triangles are decimated. In case that each triangle has a corresponding color or small texture, some colored triangles or textures are lost by the simplification. Thus, in order to preserve the appearance of the models, we need to carefully choose the vertices being decimated. In this paper, we adopt the color difference as the criterion. The idea of this second measure is as follows. Consider two vertices on 3D surfaces,  $\mathbf{v}_1$  and  $\mathbf{v}_2$  in Figure 1. While the triangles connected to  $\mathbf{v}_1$  have different color information, those of  $\mathbf{v}_2$  have similar colors. In order to preserve the appearance of the models,  $\mathbf{v}_2$  should be decimated prior to  $\mathbf{v}_1$  even when the measure  $v(i)$  may be the same for these two cases. The other measure  $c(i)$  therefore seeks to find the importance of vertices in terms of the color difference. In our current implementation, we first transform RGB values to the LUV color space. Next we find the average of each component, and use the sum of the absolute differences between the averages as the measure.

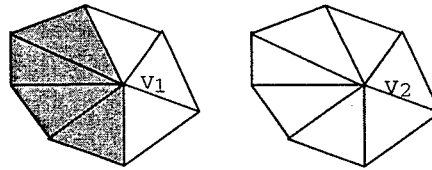


Figure 1: Left: triangles with different colors. Right: triangles with similar colors

To evaluate the overall perceptual importance of each vertex, we calculate the weighted sum of the two measures as follows.

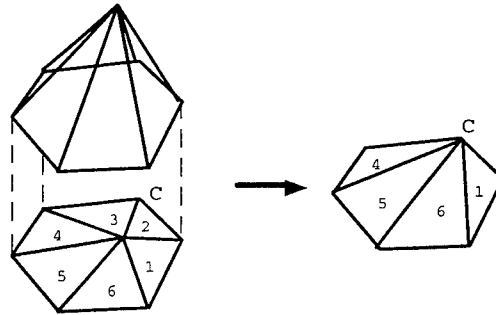
$$m(i) = \alpha v(i) + (1 - \alpha)c(i) \quad (1)$$

Vertices with large  $m(i)$  are considered more perceptually important. Therefore, they are decimated later in the encoding process, and sent earlier during the transmission process. The user can freely select the weights,  $\alpha$ , depending on the user's preference on the relative importance of geometry versus texture information. In case of 3D models without texture, only  $v(i)$  is considered.

The process of vertex decimation is applied, vertex-by-vertex, from vertices of low  $m(i)$  to vertices of high  $m(i)$ , until only irremovable vertices remain, which form the base mesh. Vertices that satisfy either one of the following two conditions are considered irremovable, so as to retain the topology and the appearance of the simplified model.

1. The triangles surrounding the vertex do not form a closed ring. Vertices on the boundaries satisfy these conditions, so this condition prevents us from decimating vertices on the boundaries.
2. The vertex has extraneous triangles connected to it. If the model has partially junctions of triangles, the vertices connected to them are not decimated.

## 2.2. Re-Triangulation and Texture Re-Mapping



C : closest vertex

**Figure 2: Re-triangulation and texture re-mapping**

Once a vertex is decimated, re-triangulation and texture re-mapping are applied to fill up the hole caused by the decimation. As we mentioned above, 3D meshes may have two types of attribute data, vertex attributes and corner attributes [12]. The corner-attributed textures are distributed at random in the texture map. The vertex attributed texture map typically looks like regular images. Thus, we need to consider both cases of corresponding the 3D geometry with the texture map. In case of the vertex attributed texture coordinates, once a vertex is decimated, the textures belonging to the triangles originally connected to the vertex can be re-mapped directly to the base left by the decimation. Since the texture coordinates of the remaining vertices will remain the same, any re-triangulation scheme produces similar results so no texture re-mapping is needed. However in case of the corner attributed texture coordinates, the re-triangulation and the re-mapping significantly affect the decimated model. We now discuss the method we use to correspond each triangle to the texture map. First, among all vertices connected to the vertex that is considered, we find the vertex that has the closest distance from the removed vertex. Then, the removed vertex is mapped to the closest vertex, which results in a triangle fan as in Figure 2. The textures owned by the triangles that are retained are then re-mapped to the new triangles. For example, in Figure 2, the decimated vertex is moved to the vertex C. The original textures of triangles 1, 4, 5 and 6 are hence mapped to the new four triangles.

### 2.3. Compression algorithm

The encoder, following the vertex decimation algorithm aforementioned, replaces the original mesh with a base mesh and a sequence of vertices with associated attribute data. Each vertex is encoded as a seven-tuple: the index of the closest vertex on the edge, the indices of the two vertices on which we start the fan triangulation, the number of triangles to traverse, the  $x, y, z$  coordinates of the decimated vertex, and two texture coordinates,  $s$  and  $t$ , for each vertex in the two triangles removed by the decimation. For 3D models without texture, texture coordinates are not necessary. Vertices in the base mesh are numbered sequentially, and each new vertex is assigned the next available index.

We encode the vertex indices using arithmetic coding. To complete the compression, we encode the vertex coordinates and texture coordinates. Since 3D objects are often well modeled as piecewise smooth regions, the vertex and texture coordinates are highly correlated. These coordinates are predictable by using the neighboring vertices. The vertex coordinates  $x, y, z$  are predicted by a linear combination of the neighboring vertices. Similarly, for the models with the vertex attributed texture coordinates, they are predicted by a linear combination of the neighboring texture coordinates.

In case of corner attributed texture coordinates, the texture coordinate of the first vertex is transmitted without any prediction and then the coordinate of the second vertex is predicted by the first coordinate. The coordinate of the third vertex is predicted by the average of the first two coordinates. The residues are quantized by a prescribed step size and then arithmetic coded.

### 2.4. Coding of Other Attribute Data

The algorithm mentioned above can be easily extended to models with other attribute data such as colors and normals. We incorporate the color and normal differences to the metric (1) in a similar way. We assign the colors and the normals by the same means of the texture re-mapping in Section 2.2. As in case of the texture, the indices and the actual data such as colors and normal vectors are transmitted separately. Once a vertex is decimated, the indices of the data assigned to the vertex are transmitted without any prediction or entropy coding. The encoder codes only the colors and the normal vectors used by the current meshes with prediction and entropy coding.

## 3. PROGRESSIVE TEXTURE CODING

As is described above, there are two types of texture correspondence to be considered, correspondence for each vertex and for each corner. In the former case, one texture image is mapped to a whole model. Since this type of texture is often as smooth as typical 2D images, we have adopted the wavelet coding algorithm in [13] to encode the texture, resulting in SNR progressive bitstreams. In the latter case, the texture map contains many triangles of various sizes. For this type of texture, we encode each triangle independently and send it to the decoder. The coding method we employ is as follows. First consider a rectangle circumscribing the triangle. We choose a rectangle with each dimension being a multiple of 8. We fill up the pixels outside the triangle with the pixels at boundary of the triangle by vertical padding followed by horizontal padding, similar to what is used in MPEG-4. Then, we compress the rectangle using discrete cosine transform (DCT), scalar quantization, zigzag scan and Huffman coding, similar to the JPEG algorithm. In our framework, only the textures required to render a current level of the model are transmitted. Hence, progressive texture coding is achieved.

## 4. EXPERIMENTAL RESULTS

### 4.1. Compression Results

We tested several 3D models in VRML format downloaded from public web sites. Table 1 shows that comparison between our algorithm and resolution progressive mode of the MPEG-4, which has 10 LOD. Our 3D representation has continuously progress resolution, i.e., the number of LOD corresponds to the number of decimated vertices plus one. In our algorithm, the base mesh of each tested model is compressed by the non-progressive codec of MPEG-4 3D coding algorithm [6].

Quantization of 10-bit resolution is applied to all vertex coordinates. From the results in Table 1, we can see that not only is our representation more “granular,” our compression efficiency is also better than the MPEG-4.

## 4.2. Progressive Streaming

We have also implemented a viewer to progressively display the 3D models coded by our algorithm. Once sufficient bits are downloaded from the server to display more detail, the new model is updated on the display. While the file is being downloaded, the user can change the viewpoint to examine the model. If the user is not interested in the model, the downloading can be stopped at any time. Figure 3 shows what a model looks like as it is being loaded through a 28.8K modem dial-up connection. This model has corner attributed textures. It can be seen that even a low level version of the model gives the user a very good idea of what the complete model would look like. Since we use the texture difference as the measure of importance for vertices, color patterns, such as the edge between white and black feathers in Figure 3(a), are nicely rendered during the whole process.

## 5. CONCLUSION

In this paper we demonstrated joint geometry/texture progressive coding of 3D models. We have created tools that code 3D files into progressive bitstreams, and a browser that allows the user to download and view these files progressively. Our viewer has been implemented and fully tested. It is available for download at <http://amp.ece.cmu.edu/>

## ACKNOWLEDGEMENTS

This work is partially supported by Japan Society of the Promotion of Science (JSPS).

## REFERENCES

- [1] H. Hoppe, “Progressive Meshes,” *Proceedings SIGGRAPH 96*, pp. 99-108. ACM SIGGRAPH, 1996.
- [2] M. Garland and P. S. Heckbert, “Surface Simplification using Quadratic Error Metrics,” *Proceedings SIGGRAPH 97*, pp. 209-216. ACM SIGGRAPH, 1997.
- [3] M. Eck, T. DeRose, T. Duchamp, H. Hoppe, M. Lounsbery, and W. Stuetzle, “Multiresolution Analysis of Arbitrary Meshes,” *Proceedings SIGGRAPH 95*, pp. 173-182, ACM SIGGRAPH 1995.
- [4] Web3D Consortium Working Groups, [http://www.web3d.org/fs\\_workinggroups.htm](http://www.web3d.org/fs_workinggroups.htm)
- [5] MPEG-4 SNHC web page, <http://www.es.com/mpeg4-snhc>
- [6] MPEG-4 SNHC, Gabriel Taubin, editor, “SNHC Verification Model 9.0 [3D Mesh Encoding],” W2301, July 1998.
- [7] R. Pajarola, and J. Rossignac, “Compressed Progressive Meshes,” Tech. Report GIT-GUV-99-05, Georgia Institute of Technology, 1999.
- [8] B. Koh and T. Chen, “Progressive VRML Browser,” IEEE Intl. Workshop on Multimedia Signal Processing, Sep 1999.
- [9] A. Khodakovsky, P. Schroder and W. Sweldens, “Progressive Geometry Compression,” preprint, California Institute of Technology, 2000.
- [10] J. Li and J. Kuo, “Progressive Coding of 3-D Graphics Models,” *Proceedings of the IEEE*, vol. 86, no. 6, June 1998
- [11] G. Taubin, J. Rossignac, “3D Geometry Compression,” no.21 in Course Notes. ACM SIGGRAPH 1999.
- [12] J.D. Foley, *Computer Graphics: Principles and Practice*, Addison-Wesley Pub Co.
- [13] A. Said and W. Pearlman, “A New, Fast, and efficient Image Codec Based on Set Partitioning in Hierarchical Trees,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243-250. 1996.



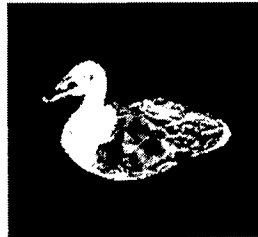
Model	#v	#t	#dv	Attributes	Proposed method (bytes)	MPEG-4 hierarchical mode (bytes)	Original VRML (KB) (*1)
Beethoven	2655	5030	2092	none	10080	13951	50
Femur	3897	7798	3824	none	13896	18832	88
Skull	10952	22104	10596	none	45076	59296	257
Triceratops	2832	5660	2762	none	10524	13831	63
Horse	11135	22258	11060	none	40299	48403	266
Duck	5013	10009	4874	texture	147811	- (*2)	605
Vase 1	5153	10044	4685	texture	126082	- (*2)	651
Vase 2	5614	10133	3338	texture	154757	- (*2)	651
Totem pole	5184	10044	3968	texture	160094	- (*2)	683
Totem pole	5184	10044	3969	color	94809	- (*2)	242
Human Face	1221	2374	1084	color	7340	- (*2)	25
Duck	5013	10009	4817	normal	80831	- (*2)	235
Vase	5614	10133	3350	normal	106222	- (*2)	258
Totem pole	5184	10044	3983	normal	92884	- (*2)	242

#v: number of vertices, #t: number of triangles, #dv: number of decimated vertices

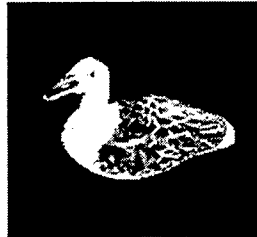
\*1: Original VRML files are gzipped and the texture models include texture files compressed by JPEG.

\*2: MPEG-4 Software does not work for these models.

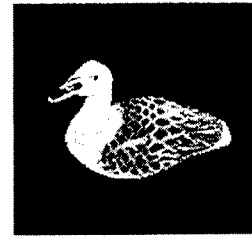
**Table 1: Comparison of total number of bytes between the proposed algorithm and MPEG-4**



(a) 8.02 sec



(b) 20.05 sec



(c) 40.10 sec

**Figure 3: Simulation of the progressive 3D models under 28.8 kbps environment**

## 3D surface digitizing and modeling development at ITRI

Wen-Jean Hsueh<sup>1</sup>

Opto-Electronics & Systems Laboratories, Industrial Technology Research Institute  
OES/ITRI-S000, Chutung, Hsinchu 310, Taiwan, ROC

### ABSTRACT

This paper gives an overview of the research and development activities in 3D surface digitizing and modeling conducted at the Industrial Technology Research Institute (ITRI) of Taiwan in the past decade. As a major technology and consulting service provider of the area, ITRI has developed 3D laser scanning digitizers ranging from low-cost compacts, industrial CAD/CAM digitizing, to large human body scanner, with in-house 3D surface modeling software to provide total solution in reverse engineering that requires processing capabilities of large number of 3D data. Based on both hardware and software technologies in scanning, merging, registration, surface fitting, reconstruction, and compression, ITRI is now exploring innovative methodologies that provide higher performances, including hardware-based correlation algorithms with advanced camera designs, animation surface model reconstruction, and optical tracking for motion capture. It is expected that the need for easy and fast high-quality 3D information in the near future will grow exponentially, at the same amazing rate as the internet and the human desire for realistic and natural images.

**Keywords:** 3D, surface digitizing, surface modeling, range image, reverse engineering

### 1. INTRODUCTION

The Industrial Technology Research Institute (ITRI) of Taiwan has been devoting to the research and development of 3D surface digitizing and modeling for almost a decade. As a major technology and consulting service provider of the area, ITRI has developed 3D laser scanning digitizers ranging from low-cost compacts, industrial CAD/CAM digitizing, to large human body scanner, with in-house 3D surface modeling software to provide total solution in reverse engineering. The core abilities to acquire 3D information from sophisticated-shaped objects and to process and manipulate its large number of 3D coordinate data have enabled the applications in reverse engineering, body scanning, and 3D animation.

### 2. METHODOLOGIES

#### 2.1 3D laser stripe digitizer

The concept of ITRI's non-contact 3D digitizer is illustrated in Figure 1. A stripe of laser is projected onto the surface of an object, and the reflected light is captured by either of the CCD cameras to avoid occlusion. Based on optical triangulation and camera calibration parameters, the 3D coordinates of the illuminated data points on the object surface can be calculated as shown in Figure 2. Integrating the digitizer head and a four-axis translation-rotation stage, as shown in Figure 5, can fully digitize a 3D object automatically. The digitizer head has a depth of field of 170 mm, and the system scans at speeds up to 3,000 points per second and precision up to 50  $\mu\text{m}$ . The digitization generates a large number of 3D coordinates from different viewing angles. Figure 4 illustrates the surface modeling process to make these unorganized 3D data useful. (See Section 2.2 for details.) Thorough discussion of the design and analysis of the digitizer can be found in [3][7][8].

<sup>1</sup> Correspondence: hsuehwj@itri.org.tw; <http://www.oes.itri.org.tw/3D>

A low-cost compact version with only one CCD camera for triangulation in PC-based applications is shown in Figures 3 and 6. It is designed so that one CCD retains the information of two fields of view of the original two-CCD design to account for occlusion issues without compromising precision and accuracy [2]. We further design the scanner to equip with a color CCD camera for building a complete color 3D model.

## 2.2 3D surface modeling

As illustrated in Figure 4, after acquiring large number of 3D coordinates from many viewing angles, registration and merging are two fundamental steps to create a useful 3D surface model. A polygon-based method for describing a complete object is proposed [6]. Multiple range images are integrated into a single polygonal, usually triangular, mesh. Registration is to align the multiple range images into the same coordinate system. Merging then removes redundant data and stitches these images to a single mesh. Our program needs no additional information from the digitizer. Figure 7 shows the reverse engineering surface modeling result of an impeller and its rapid prototyping replica.

Surface modeling from 3D laser digitizer enables the capture of detailed description of the object surface, but the large number of 3D coordinates or polygonal information hinders efficient usage and effective applications of the surface models. We proposed a sequential decimation process to reduce the number of polygons in a triangular mesh after the registration and merging of multiple range images are performed [12]. An iterative vertex decimation method is used to remove vertices with minimum re-triangulation error. To reduce the distortion resulting from polygon reduction, vertices are characterized by local geometry and topology before the re-triangulation error is evaluated. This algorithm can be applied to not only triangular meshes generated from 3D digitizer, but also general volume meshes and terrain meshes. Figure 8 demonstrates results from our polygon reduction algorithm.

One other common issue in 3D digitizing is finding the next best view to efficiently digitize the object by calculating the next best position or a working path of the sensor, which has been well addressed by peers. In developing our low-cost scanning system that has significant less degrees of freedom in movement, however, we tackle a similarly important issue of deciding the best positions of the object for efficient scanning in building an effective 3D surface model. A low-occlusion approach is proposed to find the best viewing position for scanning by considering the position of the object instead of the sensor [1]. The efficiency improves significantly by combining carefully planned working path of the sensor and optimum positions of the object.

## 3. APPLICATIONS

### 3.1 Reverse engineering

An important and original application of the 3D digitizers is in reverse engineering as already illustrated in Figure 7. We have consulted a wide range of traditional to hi-tech industries that adopts 3D digitizing as an indispensable tool for computer-aided design, manufacturing, and inspection, including applications in parts design, industrial design, tooling, sculpture, ergonomics, textile, and foot wear.

A good example is in the measurement and inspection of tires [10]. A 3D digitizing sensor is installed on a 4-axis mechanism to scan the whole tire surface in radial and circumferential directions. The laser stripe generates a section curve in the radial direction. The 3D coordinates of a series of points along the curve can be measured with an accuracy of 0.05mm. The scanned result is shown in Figure 9. The geometry of the tire can thus be evaluated, including width, diameter, circumference, arc value, and roundness. The dimensional error is always less than 0.15mm. Tread wear can also be quantified by comparing the digitized surface data to its original data, and the resolution is up to 0.1mm.

### 3.2 Body scanning

It is only natural that 3D scanning find applications in human-related measurements because of the abundant information 3D digitizers are able to provide dealing with biological and highly diversified subjects compared to the limited data from using conventional meters. As shown in Figure 10(a), our human body scanner consists of six 3D digitizers and three vertical translation axes. The range of measurement is 1900mm in height, 900mm in width, and 500mm in depth. It takes approximately 8 seconds to capture whole-body data of a 180-cm tall subject with a per-4 mm profiling, a horizontal resolution less than 2 mm, and a range resolution less than 1 mm. The body scanner satisfies the requirements for speed, safety, comfort, and efficiency, and has been used in consumer-product design and health care applications. Figure 10(b) illustrates the rendered results of scanning of a human subject [13].

### 3.3 3D animation

To obtain realistic visual effects, more and more motion picture and animation productions adopt reverse engineering technologies. 3D digitizers and motion capture systems are two important tools in this domain. Manipulating large quantity of scanned 3D data, however, is highly inefficient and difficult for animators when using 3D digitizer to build models. In the 3D modeling animation process, two issues are of major concerns: fast surface reconstruction and easy surface structure manipulation. Fast surface reconstruction saves time and thus cost. Easy surface structure manipulation asks for continuity maintenance and flexible orientation of coordinate system for merged surface reconstruction. We proposed a fast surface reconstruction pipeline of the 3D digitizer and used a motion capture system to integrate surface modeling technologies [9]. Figure 11 shows the results after clustering, surface reconstruction, and motion capture integrated animation rendering.

## 4. OUTLOOK

It is expected that the need for easy and fast high-quality 3D information in the near future will grow exponentially, at the same amazing rate as the internet and the human desire for realistic and natural images. 3D imaging can be achieved from many sources, including active scanning for model-based 3D and passive capture for vision-based 3D. Either one has its advantages, and integration at some level is expected to be a certain route towards advancement of the technology. The requirement of performances will certainly go in the direction of real-time, dynamic, high-resolution, and high-accuracy 3D imaging.

Based on both hardware and software technologies in scanning, merging, registration, surface fitting, reconstruction, and compression, ITRI is now exploring innovative methodologies that provide higher performances. A project involving hardware-based speckle image correlation algorithms with advanced camera designs [4][11] wishes to provide high-speed 3D imaging capabilities. Effort in animation application will continue with integration of capabilities in high-speed tracking for motion capture. Acquiring 3D information from vision-based cameras will also be pursued. Topics in creating natural 3D color will be explored. Integration among 3D input, processing, display, and human interface technologies to spawn more innovative ideas in human-centered technologies is underway.

## 5. ACKNOWLEDGEMENTS

The R&D activities that lead to the technologies described above were made possible by continuous funding and support from the Ministry of Economic Affairs of Taiwan, ROC for the last eight years and continuing. The author would like to acknowledge outstanding contributions made by all members of the Optical Inspection Department at the Opto-Electronics & Systems Labs of ITRI that made the work successful, with special thanks to Mr. Hsien-Chang Lin, Mr. Chia-Chen Chen, and Dr. Bor-Tow Chen for their generous help and suggestions to this article.

## 6. REFERENCES

1. Chen, B.-T., W.-S. Lou, C.-C. Chen, and H.-C. Lin, "A 3D scanning system based on low-occlusion approach," *2<sup>nd</sup> 3DIM Conf. 3-D Digital Imaging and Modeling*, 506-515, Ottawa, Canada, 1999.
2. Chen, B.-T., W.-S. Lou, C.-C. Chen, and H.-C. Lin, "Low-cost 3D range finder system," *SPIE Proc. Input/Output and Imaging Technologies*, **3422**:99-107, Taipei, Taiwan, 1998.
3. Chen, B.-T., W.-S. Lou, C.-C. Chen, and H.-C. Lin, "3D digitizer: method and analysis," *10<sup>th</sup> IPPR Conf. Computer Vision, Graphics and Image Processing*, 406-412, Taichung, Taiwan, 1997.
4. Hsueh, W.-J. and D. P. Hart, "Real-time 3D topography by speckle image correlation," *SPIE Proc. Input/Output and Imaging Technologies*, **3422**:108-112, Taipei, Taiwan, 1998.
5. Hsueh, W.-J. and E. K. Antonsson, "Automatic high-resolution optoelectronic photogrammetric 3D surface geometry acquisition system," *Machine Vision and Applications*, **10**(3):98-113, 1997.
6. Lee, C.-M., H.-M. Tsai, C.-C. Chen, and H.-C. Lin, "Multiple range views preprocess – alignment and merging," *12<sup>th</sup> IPPR Conf. Computer Vision, Graphics and Image Processing*, 355-359, Taipei, Taiwan, 1999.
7. Lee, C.-S. and Y.-S. Wen, "Non-contact laser stripe 3D digitizer & the surface reconstruction technique for reverse engineering," *Asia-Pacific Symp. Instrumentation*, 126-132, Huangshan City, China, 1997.
8. Lee, C.-S., C.-M. Lee, and M.-W. Lin, "Laser stripe 3D digitizer and application," *12<sup>th</sup> IPPR Conf. Computer Vision, Graphics and Image Processing*, 337-343, Taipei, Taiwan, 1999.
9. Liang, C.-C. and C.-C. Chen, "Fast animation surface model reconstruction for 3D laser scanning," *12<sup>th</sup> IPPR Conf. Computer Vision, Graphics and Image Processing*, 360-364, Taipei, Taiwan, 1999.
10. Lin, M.-W. and W.-S. Lou, "Application of 3D digitizer on inspecting tire geometry," *12<sup>th</sup> IPPR Conf. Computer Vision, Graphics and Image Processing*, 344-349, Taipei, Taiwan, 1999.
11. Rohaly, J. and D. P. Hart, "High resolution, ultra fast 3-D imaging," *SPIE Proc. 3D Image Capture & Applications*, **3958**, San Jose, California, USA, 2000.
12. Tsai, H.-M., C.-C. Chen, and H.-C. Lin, "Simplification of 3D triangular meshes," *11<sup>th</sup> IPPR Conf. Computer Vision, Graphics and Image Processing*, Taipei, Taiwan, 1998.
13. Yang, Y.-X., W.-S. Lou, M.-T. Su, B.-H. Wang, and M.-W. Lin, "Whole body scanner: a true meter for body scale within few seconds," *12<sup>th</sup> IPPR Conf. Computer Vision, Graphics and Image Processing*, 350-354, Taipei, Taiwan, 1999.

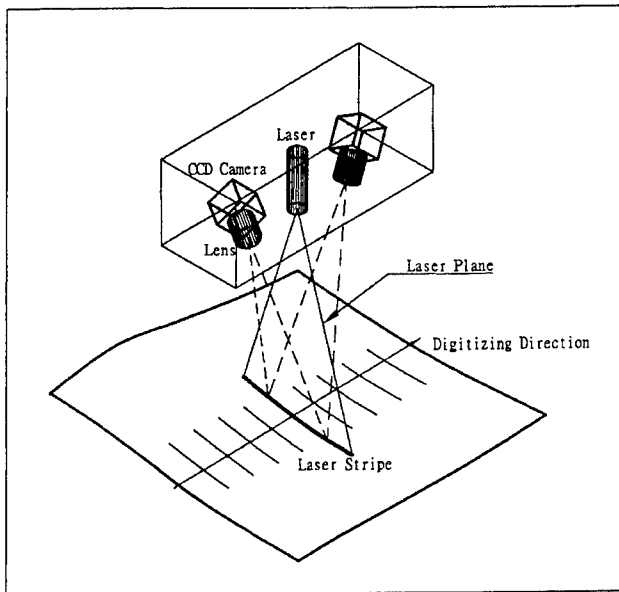


Figure 1. 3D laser stripe digitizer schema.

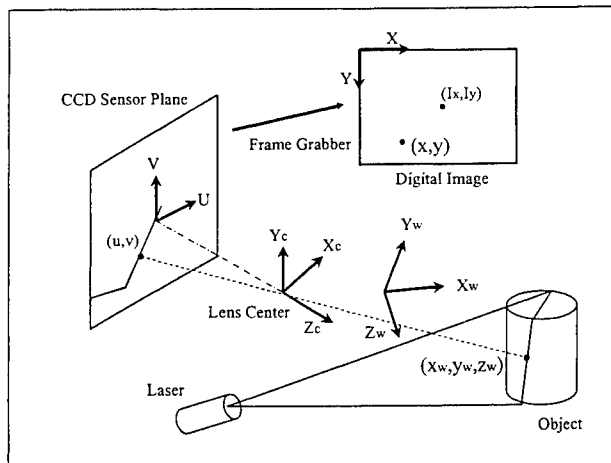


Figure 2. 3D image formation.

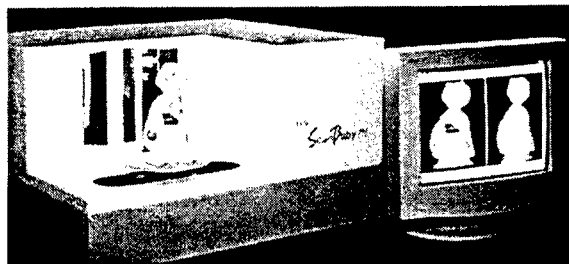


Figure 3. ITRI low-cost 3D scanner.

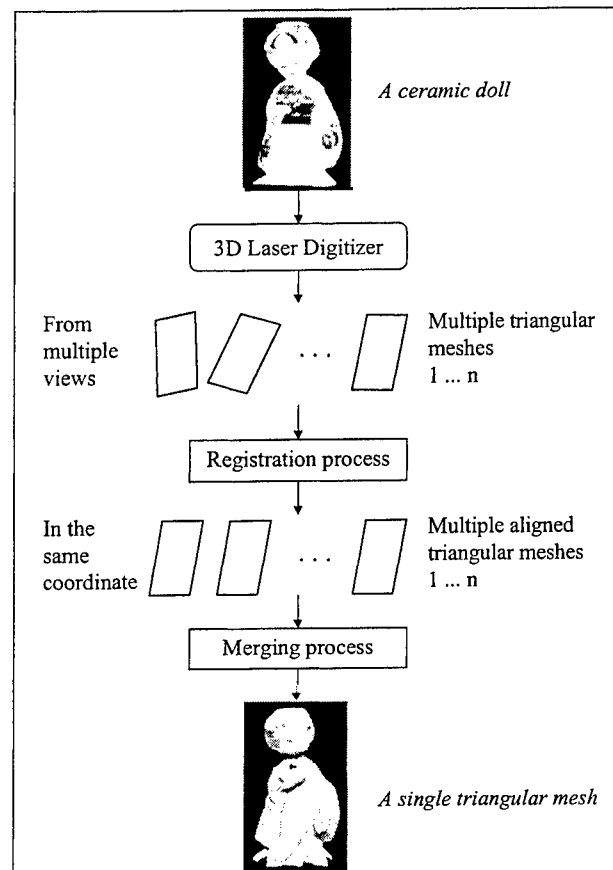


Figure 4. Procedure of surface model integration.

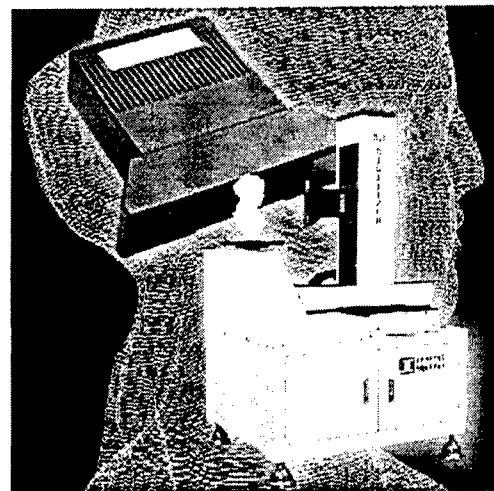


Figure 5. ITRI 3D laser digitizer.

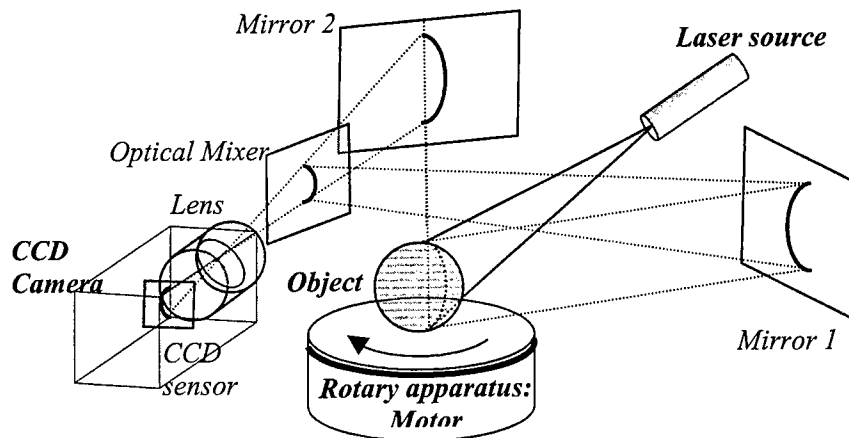


Figure 6. Low-cost scanner schema.

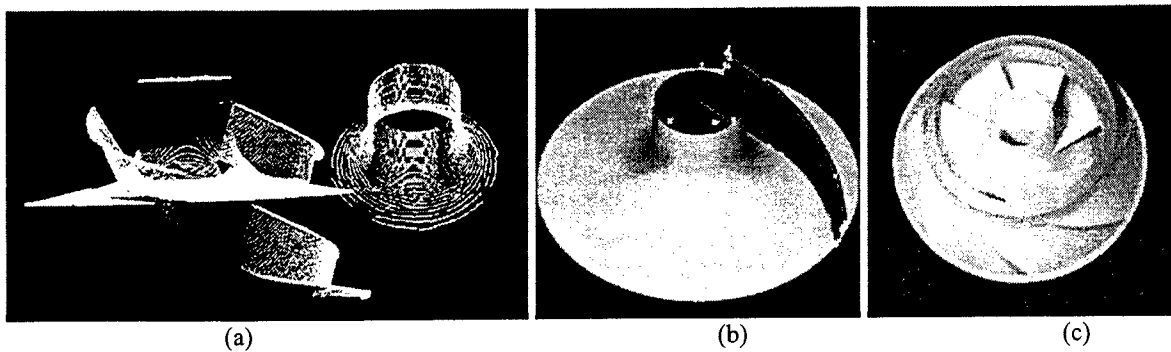


Figure 7. Impeller surface modeling. (a) Multi-view range data, (b) fitted impeller vane surfaces, (c) rapid prototyping.

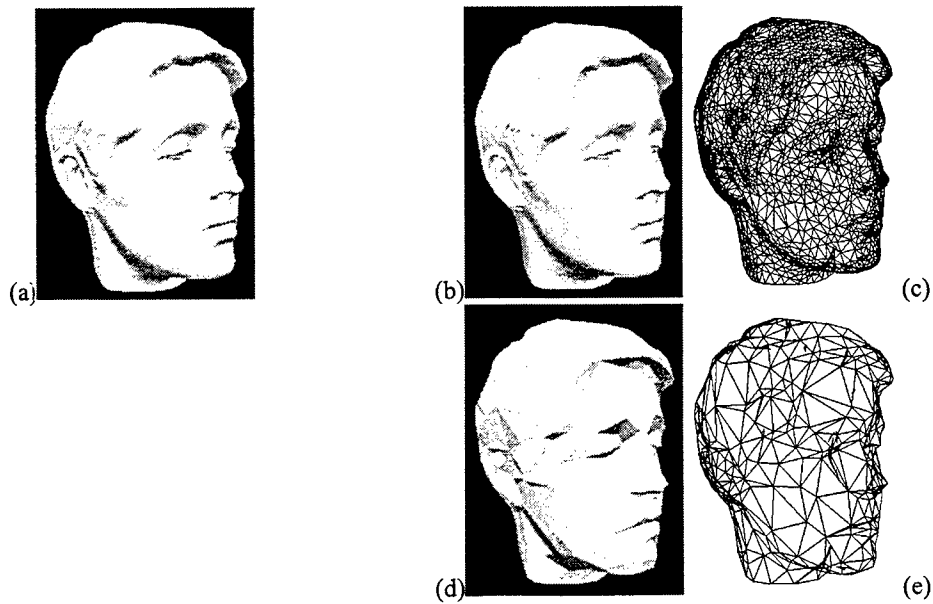
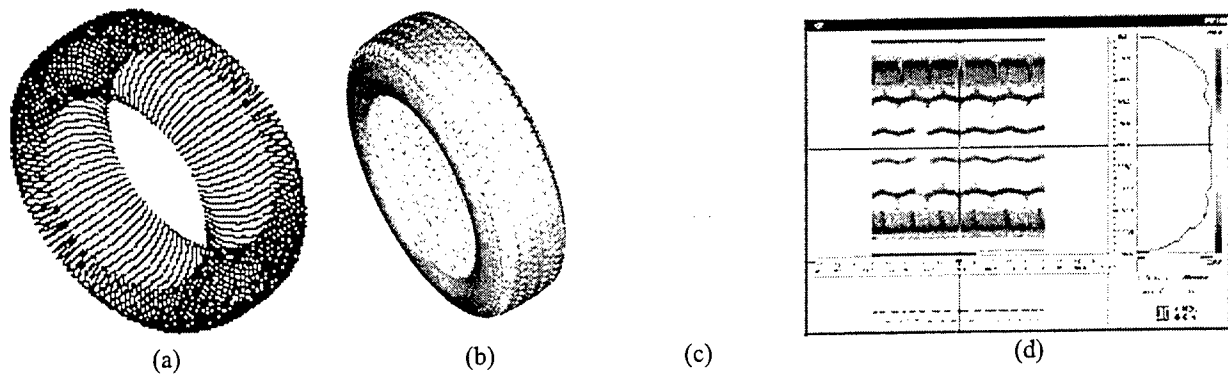
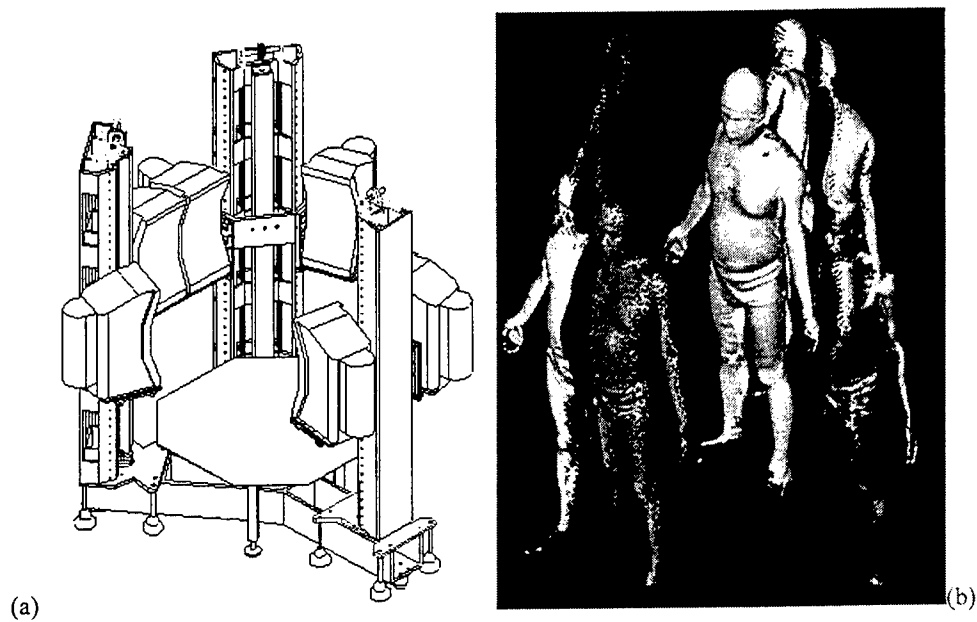


Figure 8. Polygon reduction of a human head model. (a) Original flat-shaded model from 3D digitizer (43,359 triangles), (b) 75% polygon reduction (10,795 triangles) flat-shaded, (c) wire-frame, (d) 97% polygon reduction (1,262 triangles) flat-shaded, (e) wire-frame.



**Figure 9.** Application in reverse engineering for tires. (a) Measured 3D profile, (b) shaded image of tire, (c) radial section profile, (d) wear inspection.



**Figure 10.** ITRI whole body scanner. (a) System schema, (b) multi-view merging and rendering of measured human body 3D data.



**Figure 11.** Application in 3D animation. (a) Clusters in a facial model, (b) surface patches, (c) rendered images of facial animation.



# A Surface Reconstruction Technique based on 3-D Triangulation Enhancement

I-Cheng Chang<sup>1</sup>, Bor-Tow Chen, Kun-Jiang Hsieh, Wen-Jean Hsueh and Hsien-Chang Lin

S200, Optical Inspection Department, Opto-Electronics & Systems Laboratories,  
Industrial Technology Research Institute, Chutung 310, Taiwan, R.O.C.

## ABSTRACT

3-D reconstruction technique plays an important role in the applications for 3-D data acquisition, such as medical diagnosis, animation and virtual reality. Moreover, the 3-D triangulation process is one of the most important parts while reconstructing the smooth surface of a 3D object. The essential of 3-D triangulation is to find the intersection of the rays emitting from the correlated points on each image pair, but the emitting rays always don't intersect with each other owing to the error in process. So the obtained 3-D point is an approximated value and makes the reconstructed surface uneven. In the paper, we propose an triangulation enhancement method, which reduces the perturbation in the reconstructed data and filter out the error caused by spurious vectors in the process of correlation.

**Keywords:** 3-D geometry, correlation vector field, 3-D triangulation, multiply factor, 3-D stereo imaging system.

## 1. INTRODUCTION

The reconstruction of 3D object from stereo images has been widely investigated for various applications, such as medical diagnosis, virtual reality, and animation, etc. In the processes of reconstruction, 3D data are mainly determined by data correlation between two image frames and 3D triangulation from two correlated points. The correlation process is adopted to calculate the correspondence of data points of two image frames to generate correlated vector fields. According to the relation provided by the vector field, the triangulation process will produce 3D coordinates of an object. The concept of the technique is to restore the 3-D coordinates of an object by using the point correlation on two image frames and the extrinsic/intrinsic parameters[1]. The extrinsic parameters depict the location and orientation of the camera reference frame with respect to a world reference frame, whereas the intrinsic parameters describe the connection of the pixel coordinates of an image point and the corresponding coordinates in the camera reference frame. Whether the parameters are known or not, Trucco[1] classified the reconstruction problems into three cases: (1) intrinsic and extrinsic parameters are known, (2) only intrinsic parameters are known, and (3) no information on parameters. However, due to the error of correlation and calibration processes, it is still impossible to get a totally correct solution even if in case of (1). The phenomena of case(1) are the fundamental issue we should cope with. Our research aims to find an enhanced triangulation method, which could filter out the spurious vectors in the correlation vector field and reduce the 3-D error.

In Section 2, we describe the 3-D stereo imaging system based on the direct triangulation and an enhancement triangulation is proposed in section 3. Section 4 illustrates the experimental results and finally section 5 gives the conclusions.

## 2. 3-D STEREO IMAGING SYSTEM BASED ON DIRECT TRIANGULATION

### 2.1 Stereo imaging system

In the paper, a 3-D stereo imaging system with laser speckle projection is proposed to capture the stereo images and reconstruct the shape of a 3-D object(see Figure 1).

---

<sup>1</sup> Correspondence: Email: [h880647@itri.org.tw](mailto:h880647@itri.org.tw); Tel: 886-3-5918414; Fax: 886-3-5829781

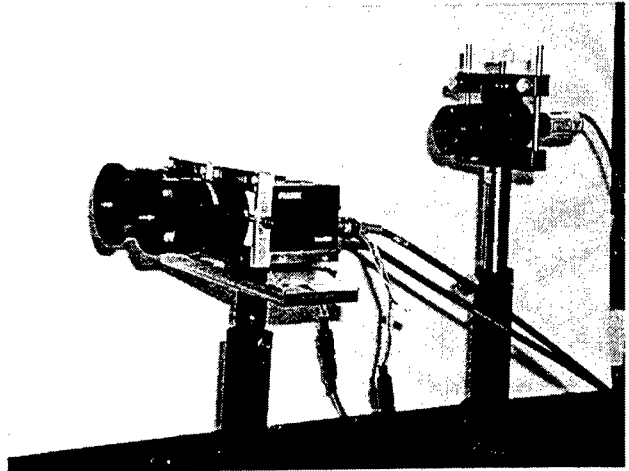
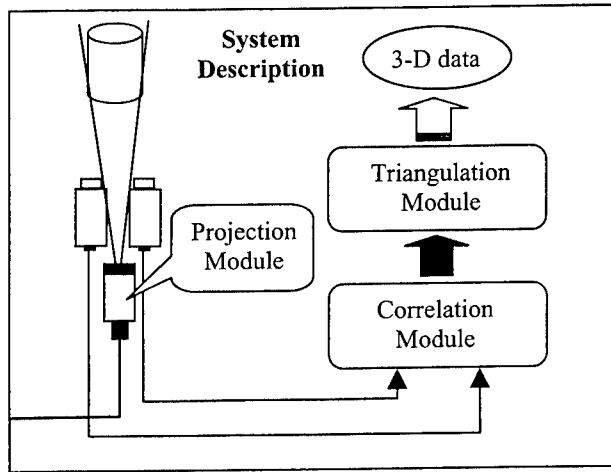


Figure 1 The 3-D stereo imaging system with laser speckle projection.

In the aspect of processing, a 3-D stereo imaging system contains three major modules: (1) projection module, (2) correlation module and (3) triangulation module. In the projection module, the randomized patterns are produced by letting laser beam pass through a diffuser and projected on the object surface, which provides the reference information for the correlation module. There are some criteria about how to select an appropriate speckle pattern for better correlation result([2]). Then, in correlation module, we compute the correlation by using the relation between the two image frames and build the correlation vector field, which describes the correlation of the points on left and right image frames. In order to get high speed processing, we adopt the method of compressed image correlation proposed by Hart([3], [4] and[5]). The method has been successfully applied to track the flow vectors in PIV images.

Finally, the triangulation module is adopted to calculate the 3-D coordinates from the vector field and reconstructs the 3-D object. The essential of triangulation module is to find the intersection of the rays emitting from the pair of correlated points. In ideal case, the 3-D rays corresponding to the correlation points on the image frames should intersect with each other at a point,  $P''$ (see Figure 2). However, in practical situation, the two rays will be skew in 3-D space without intersection. The triangulation error comes mainly from two error sources: (1) correlation error and (2) CCD calibration error. We identify three conditions by considering the two error sources:

- The correlation and CCD calibration is perfect and error free.  
The de-projection rays( $L$  and  $R''$ ) intersect at point  $P''$ (see Figure 2).
- The correlation is perfect but with CCD calibration error.  
The correlation pair is still  $(P_L, P'_R)$ , however, the de-projection ray  $R''$  corresponding to  $P'_R$  moves to  $R'$  because of the calibration error(see Figure 2). The triangulation point  $P'$  is estimated from the shortest path between both rays.
- Both correlation and CCD calibration induce error.

This is the most common condition. The correlated point in the right image frame shifts from  $P'_R$  to  $P_R$ , moreover, the de-projected rays is  $R$  and the triangulation point becomes  $P$ (see Figure 2).

In the paper, our system is based on the third condition since the the results of the correlation and calibration processes should bring error in real world. The computation of triangulation  $P$  is as following:

(1) Let ray  $L$  be  $\alpha P_L$  and ray  $R$  be  $T + \beta R^T P_R$ , then the vector  $w$ , which is orthogonal to both rays  $L$  and  $R$ , is described as  $w = P_L \times R^T P_R$ . (Eq.1)

(2) Solve the equation

$$\alpha P_L - \beta R^T + \gamma w = T. \quad (\text{Eq.2})$$

Solving Eq.2 will determine the end points of the shortest path(s) between two rays, and the triangulation point  $P$  is the midpoint of the path s.

The correlation error dominates the 3-D triangulation error because the 2-D errors are enhanced while 2-D points in image frames are converted to 3-D rays. In section 2.2, the 3-D error caused by correlation error will be analyzed and discussed.

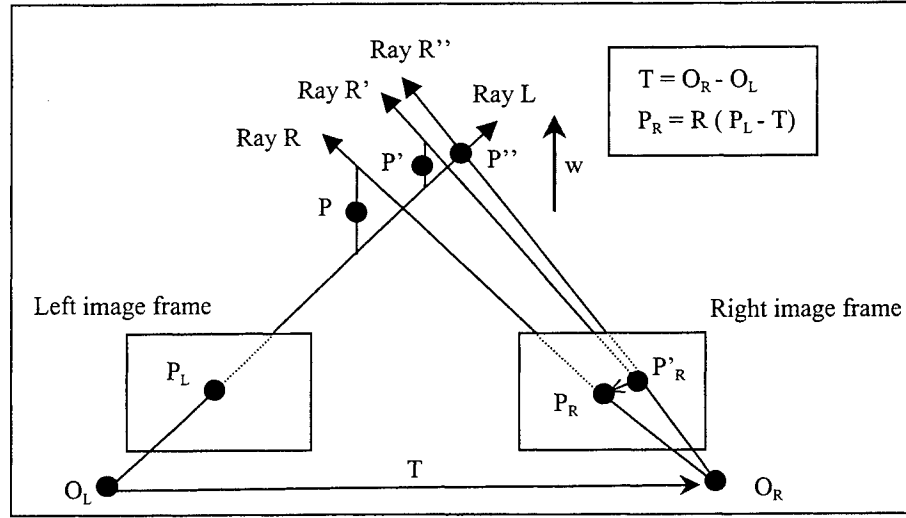


Figure 2 Three conditions in triangulation process.

## 2.2 Relation between 2-D shift and 3-D displacement

Many issues affect the correctness of correlation, for example, the quality of the spackle projection, the slope of object surface, the CCD inherent disturbance, etc. We make an experiment to investigate and measure the effect on 3-D space when the correctness of correlation vector is corrupted by the issues:

Step 1. Take a set of CCD parameters obtained in the calibration process.

Step 2. Grid the left image frame and use the position of cross-intersection as the left point set( $S_L$ ).

Step 3. Assume the 3-D surface is  $z=0$  and de-project  $S_L$  the 3-D point set( $S_{3-D}$ ) by using the CCD parameters.

Step 4. Project  $S_{3-D}$  to right image frame and get the right point set( $S_R$ ).

Step 5. Shift the right point set in x- or y- direction as the correlation error to calculate the influence on the coordinates of the triangulation point.

Figure 3 shows the relationship between mean/STD and x/y shift. Figure 3(a) illustrates that the 3-D displacement order is around  $10^{-3}$  in the shift of y-direction and Figure 3(b) represents that the shift in x-direction has much displacement, the order is around  $10^{-1}$ . Besides, since the STD is low in both, it depicts that the point on the image frame with the same shift has similar effect on 3-D displacement.

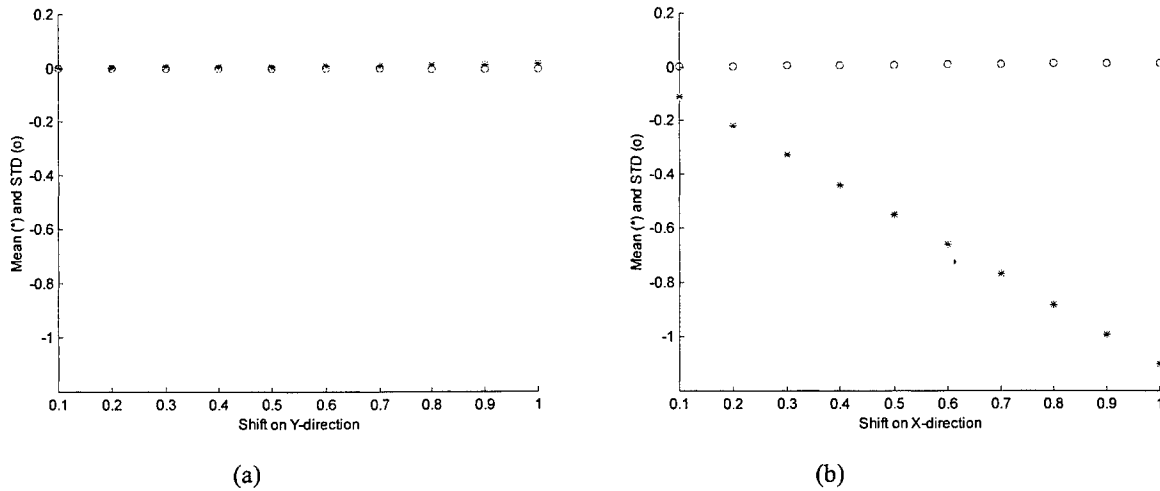


Figure 3 Mean and STD on Z-direction direction.

From the observation, we conclude that the depth-shift ratio is 0.8~1.1 mm/pixel in x-direction, in other words, the z coordinate will shift about 1mm if the correction vector has pixel-level error. Of course, the depth resolution is dependent on the photogeometry of the stereo system, e.g. the distance between CCDs. If the distance between CCDs increases, then the depth resolution will increase. But it is unreasonable to enlarge the distance unlimitedly. In our system, the distance is approximately 60 mm.

### 3. THE ENHANCEMENT TRIANGULATION

The system based on direct triangulation method faces two major problems:

- (1) The z-coordinates of the reconstruction surface are sensitive to the correlation error.
- (2) Some spurious vector still survive after Correlation Error Correction(CEC) in the correlation process.

Problem (1) will be discussed in section 3.1 and section 3.2 gives an solution to problem(2).

#### 3.1 Vector Mapping Method

Since the stereo system has two CCDs, the direct triangulation will enlarge the error while the 2-D correlated pair is converted to 3-D space rays. Little shift of correlation vectors induces larger 3-D depth error. A vector mapping method is proposed to directly convert the 2-D pixel coordinate  $(x, y)$  to 3-D space coordinate  $(X, Y, Z)$  by using three scale factors  $(K_x, K_y, K_z)$ . This will reduce the 3-D perturbation caused ion from . The central part of the method is to compute the scale factors.

Figure 4 illustrates the computation of  $K_x$  and  $K_y$ . Select a observation space where the object will be located, then calculate the scale factors  $K_x$  and  $K_y$  by using the relationship between  $(X_2-X_1, Y_2-Y_1)$  and  $(x_2-x_1, y_2-y_1)$ .

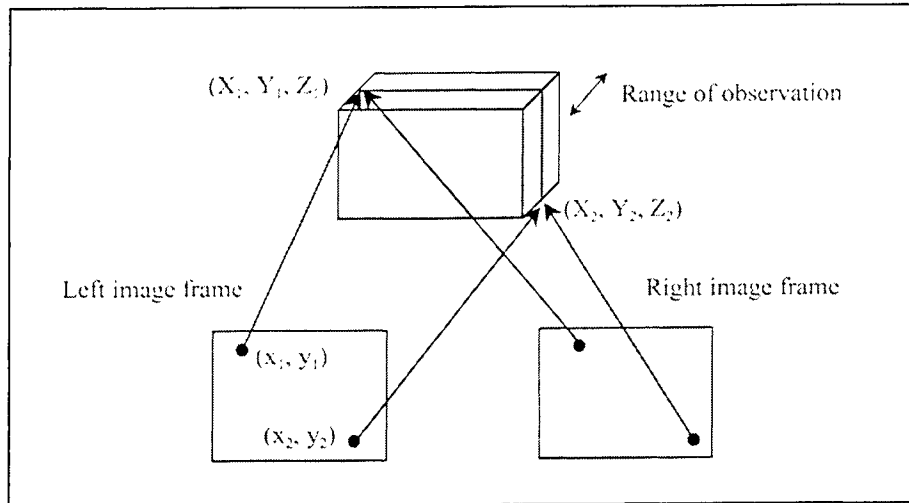


Figure 4 Computation of  $K_x$  and  $K_y$ .

The length of the correlation vector is variant when a 3-D point moves with different distance from CCDs. According to the relation of the vector mean length related to the varied distance between the 3-D point and CCDs,  $K_z$  is computed from the ratio of the mean length and distance. Surely, the value of  $K_z$  should vary with the distance, however,  $K_z$  will approximate a constant if the z-depth of the point is in a certain range. Here, we use a template to calibrate  $K_z$  (see Figure 5). Firstly, we place the template at the position at  $Z=0$ , and compute the mean length( $len(V_0)$ ) of the vectors in correlation vector field. Then we move the template to another position  $Z = t$ , also compute the mean length( $len(V_t)$ ).  $K_z$  related to position  $Z = t$  is calculated from  $t / (len(V_t) - len(V_0))$ .

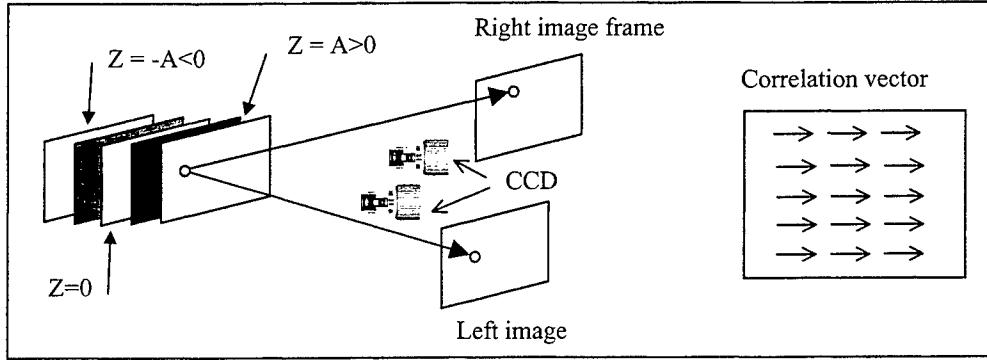


Figure 5 Calibration of  $K_z$ .

The relationship between 2-D pixel coordinate  $(x, y)$  and 3-D space coordinate  $(X, Y, Z)$  is formulated as

$$\begin{cases} X = X_s + kx * x \\ Y = Y_s + ky * y \\ Z = K_z * (len(Vx) - len(Vx_0)) \end{cases} \quad (\text{Eq. 3})$$

where  $X_s$  and  $Y_s$  are the initial position of 3-D space.

### 3.2 Distance Map

The 3-D rays of the correlated point pair often doesn't intersect with each other because of the correlation and calibration errors. The distance between two rays gives a direction about how to filter the spurious vectors. Figure 6 shows the experiment of computing the shape object of half-sphere.

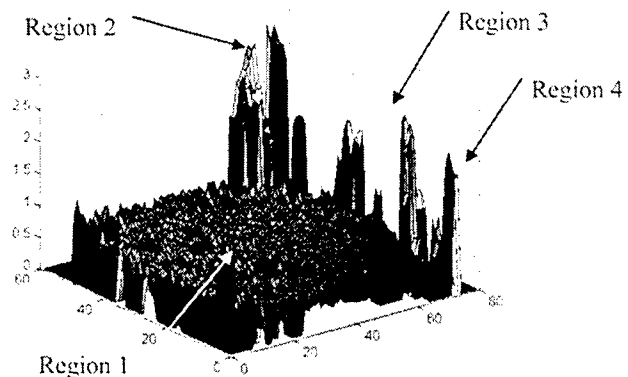
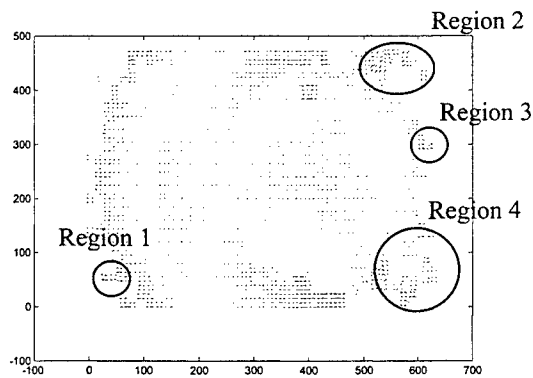
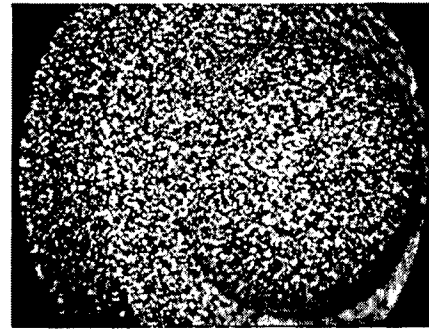
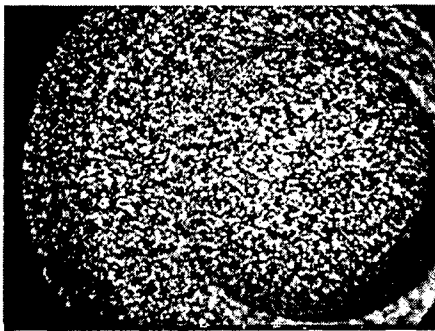


Figure 6 Correlation vector field and distance map of the half sphere.

Figure 6(a) and 3(b) are the left and right speckle images of the half sphere. The correlation vector field corresponding to the two images is shown in Figure 6(c). According to the smoothness criteria, we find that maybe some spurious vectors occur in region1 to region4, where the vectors are obviously in the different direction. However, the vectors could possibly locate on the edge or peak of the object. We cannot make sure that they are exactly spurious vectors or not even if they pass the CEC process in the correlation module. But from the distance map(see Figure 6(d)) corresponding to the vector field, the related areas have significant high distance. In other words, there is a proportional relationship between distance map and correlation vector field, i.e., some position with large value in distance map represents the corresponding vector is spurious vector. This phenomenon provides a useful indication to identify the spurious vectors. We can filter out the spurious vector by observing the corresponding distance in the map.

### 3.3 Procedures Related to Enhancement Triangulation

Figure 7 illustrates the related procedures of triangulation enhancement. The method includes the generation of distance map exploited to filter out the spurious vectors and the calibration of multiple factors ( $K_x$ ,  $K_y$ ,  $K_z$ ) applied to compute the 3-D coordinates of object.

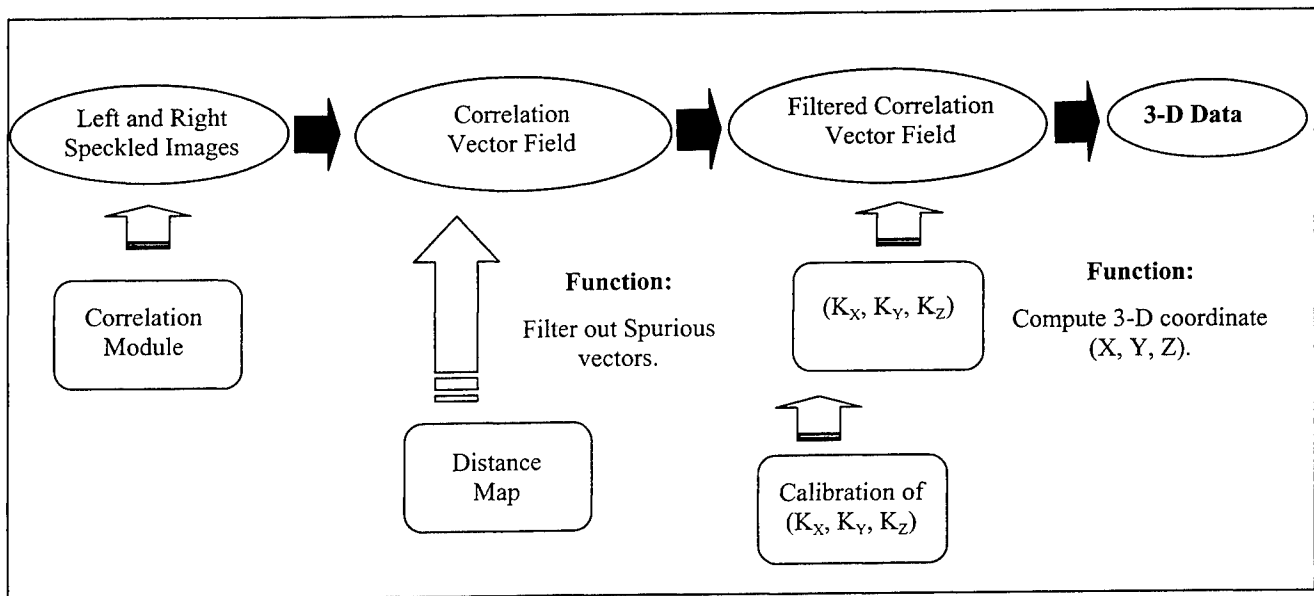


Figure 7 Stages of triangulation enhancement.

## 4. EXPERIMENTAL RESULTS

In the section, we compare the experimental results from direct triangulation and triangulation enhancement separately. The distance between two CCDs is 65 mm and the object is 550 mm from the CCDs. Figure 8 displays the left and right speckle projection images captured from the stereo system.

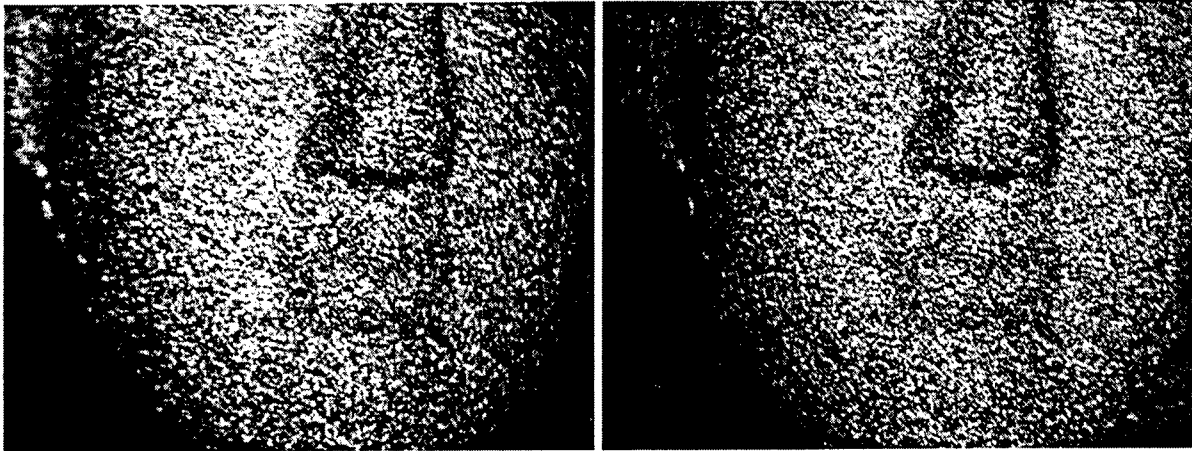


Figure 8 Left and right images of Venus Head.

Figure 9 is the result by using the direct triangulation method and Figure 10 shows the reconstructed surface through the method of triangulation enhancement.

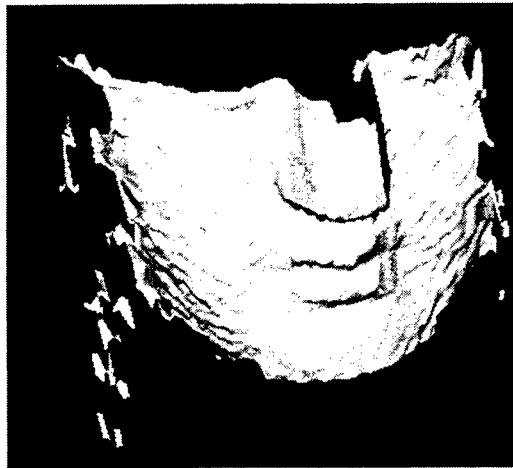


Figure 9 The reconstructed object from direct triangulation.

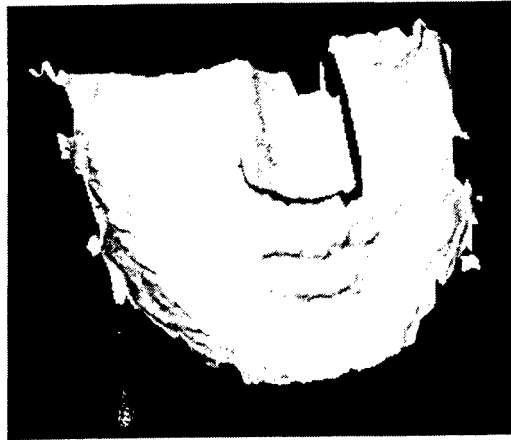


Figure 10 The reconstructed object from triangulation enhancement.

Compared Figure 9 with Figure 10, the surface points with large error are filtered after the enhancement, especially on the right cheek. Besides, the surface in Figure 10 is smoother than Figure 9, and the difference is particularly obvious on the left cheek. The reason is that the histogram of the spackled image is not very the same(see Figure 8), so the correlation errors in the part of left cheek are larger than other area of the face. The direct triangulation will enlarge the influence of the correlation error while converting 2-D image pairs to 3-D points, however, the triangulation enhancement has smaller perturbation in 3-D surface since it just reflects the correlation error.

## 5. CONCLUSIONS

In the paper, we propose a triangulation enhancement in a stereo imaging system. The proposed method gets smoother surface than direct triangulation by reducing 3-d error through vector mapping method, moreover, it filters out the spurious errors by exploiting the distance map. By observing the experimental results, we conclude that over 85% of the 3-D errors are influenced by correlation error. Therefore, the correlation algorithm on sub-pixel level should be improved to increase the correlation accuracy. Furthermore, the depth resolution is another important consideration; in the other words, the reconstructed surface will achieve better quality if the depth-shift ratio becomes lower. Our future work will focus on how to improve the accuracy of the correlation algorithm and increase the depth resolution.

## REFERENCES

1. Emanuele Trucco and alessandro Verri, "Introductory techniques for 3-D computer vision," Prentice-Hall, 1998
2. Bor-Tow Chen, Kun-Jiang Hsieh and I-Cheng Chang, "Report for the project of quantified video rate three-dimensional imaging," Technique Report, 2000.
3. Douglas P. Hart, "High-speed PIV analysis using compressed image correlation", Journal of Fluids Engineering 120, pp. 463-470, 1998.
4. Janos Rohaly and Douglas P. Hart, "High resolution, ultra fast 3-D imaging", Proceedings of SPIE, Vol. 3958, pp. 2-10, January 2000.
5. Wen-Jean Hsueh and Douglas P. Hart, "Real-time 3-D topography by speckle image correlation", Proceedings of SPIE, Vol. 3422, pp.108-112, July 1998.



# Facial Model Estimation (FME) Algorithms Using Stereo/Mono Image Sequence

Tsang-Gang Lin<sup>a</sup> and Chung J. Kuo<sup>b</sup>

<sup>a</sup>Opto-Electronics & Systems Laboratories, Industrial Technology Research Institute,  
Chutung, Taiwan 31040

<sup>b</sup>Institute of Communication Engineering, National Chung Cheng University,  
Chiayi, Taiwan 62107

## ABSTRACT

Facial model generation is an important issue in the model-based applications, such as MPEG-4 and the virtual reality. An effective and precise construction algorithm of 3D facial model from 2D images should be available for practical applications. To generate facial model usually requires stereoscopic view of the face in the pre-processing stage. Although facial model can be successfully estimated from two stereo facial images, the occlusion effect and imprecise location of the feature point prohibit us from obtaining an accurate facial model. In this paper, we proposed several facial model estimation (FME) algorithms to find the precise facial model from a stereo or mono image sequence. The information of head movement, which is recorded in the image sequence, in the temporal domain is utilized for the facial model estimation. Even though the *a priori* information about the 3D position of the head with respect to the camera and the rotation axis and angle of the head's movement are unknown, an accurate facial model (within 7.21% error) can still be obtained by our schemes. In addition, our schemes do not require the precise camera parameters and avoid the tedious camera calibration such that the facial model generation is easily achieved.

**Keywords:** Facial model, Model generation, Model-based coding, Facial image sequence

## 1. INTRODUCTION

Very low bit rate coding of (stereo) video signals is very demanding due to the advent of visual communications and limit of communication channel.<sup>1</sup> For the applications of very low bit rate video coding, the (face and body) model-based coding defined in MPEG-4 standard<sup>2</sup> is a very promising scheme. It is known that facial image is the most important type of images within the video signals. Therefore, most model-based coding schemes are thus concentrated in facial images.<sup>3</sup>

In facial model coding, 3D model of a human face is first synthesized. 3D facial model can be extracted from stereo facial images from, at least, two perspective views which are usually obtained by using two cameras separated by a distance  $d_s$ .<sup>4</sup> The two cameras should be calibrated and their parameters should be obtained in order to calculate the facial model. Since camera calibration and camera parameters extraction are tedious works, the 3D facial model is thus difficult to obtain.<sup>5</sup> In addition, some occlusion effects also exist within a pair of stereo image. In this case, the position of the occluded feature points cannot be estimated precisely.

Several algorithms for adapting facial model have been proposed. A facial model is adapted manually by Aizawa and Saito.<sup>6</sup> Some recent research<sup>7</sup> show that a facial model can be obtained from frontal and side view images, and still some<sup>8</sup> show that the facial model can be derived from stereo image pair. In estimation of the 3D position from the corresponding points in the left- and right-view image frames, the mismatch due to the mis-correspondence creates large errors in the facial model. To solve these problems, we propose a facial model updating algorithm by using the redundancies which come from the motion and disparity of the stereo image sequence.

Practically most image sequences consist of mono-scopic but not stereo view. Here, we proposed other two algorithms to extract the facial model from a mono image sequence. The first one extract the model parameters from two consecutive image frames (at two different time), while the second one updates the facial model during the course of time in mono image sequence.

---

Further author information: (Send correspondence to Chung J. Kuo)

Chung J. Kuo: E-mail: kuo@ee.ccu.edu.tw

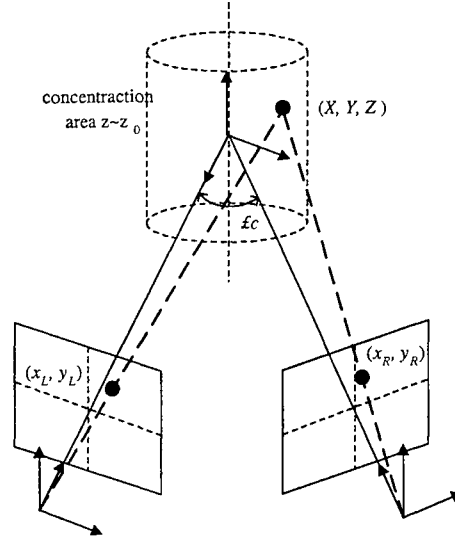


Figure 1. The geometry of the stereo imaging system

## 2. 3D HUMAN FACE MODELING

### 2.1. 3D Facial Modeling through Stereo Images

The estimation of 3D position in a stereo imaging system is accurate given the well-calibrated parameters, but the calibration of stereo imaging system is quite complicated. Chung and Nagata<sup>9</sup> proposed a camera model in which only the focal length  $f$ , the convergence angle  $\theta$  (the angle between the optical axes of two cameras), and the baseline distance  $D$  (the distance between the centers of two cameras) are necessary. These parameters can be obtained by direct measurement or found in the camera data sheet.

As shown in the figure 1, the distances from both camera centers to the world coordinate origin are the same, and the value of  $z_0$  can be simply driven using its internal parameters as

$$z_0 = \frac{D/2}{\sin(\theta/2)} \quad (1)$$

If a point in the concentration area ( $z \sim z_0$ ) is projected into the left and right image planes and their image plane coordinates are respectively  $(x_L, y_L)$  and  $(x_R, y_R)$ , the coordinates of the point in the left camera frame are approximately determined as

$$X = \frac{z_0}{f} x_L \quad (2)$$

$$Y = \frac{z_0}{f} y_L \quad (3)$$

$$Z = \frac{z_0}{f} \frac{x_L \cos \theta - x_R}{\sin \theta} \quad (4)$$

### 2.2. 3D Facial Modeling through Single Front-view Image

The essential concepts of lateral parameter estimation can be described as follows. If we want to estimate the depth of mouth,  $\theta$ , for one unknown person, we must first obtain the observation,  $Y$ , such as the width of mouth and several related parameters from his frontal face. The raw value of mouth width is converted into standardized score,  $Z_Y$ . Suppose this score is  $1.1SD + \text{mean}$  in the population of our database, what will be the possible z-score of the lateral mouth depth? To solve such problem, we must obtain the conditional probability of  $Z_\theta$  given  $Z_Y = 1.1SD$ ,

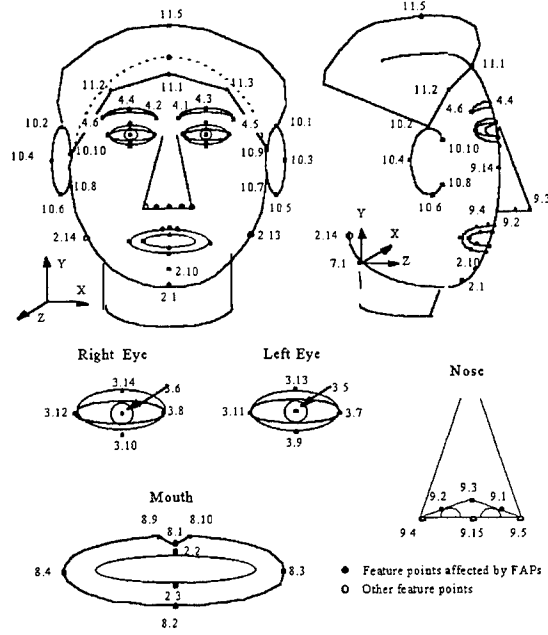


Figure 2. The feature points used in this study.

and estimate  $Z_\theta$  by the previous discussed rules. We may obtain a set of estimates for  $Z_\theta$ . Therefore, a criterion for selecting one as the best estimate must be made. Finally, we convert the estimated  $Z_\theta$  into real value in metric, and thus we obtain the depth of mouth.

### 3. FACIAL MODEL ESTIMATION THROUGH STEREO IMAGE SEQUENCE

Feature extraction is the preliminary procedure for the advanced processing. We adopt the scheme including the template matching and correlation techniques similar to Nguyen and Huang.<sup>10</sup> The feature point used in this study is a sub-set shown in Fig. 2. Once all the feature points in the face are located from a stereo image pair, the facial model can be estimated by calculating the disparity between the feature points in the left and right image frame. Theoretically, in stereo image sequence, stereo matching is only necessary for the initial pair of frames. For the subsequent pairs, stereo correspondence can be found from the temporal motion vectors and stereo matching would be necessary only for the features that newly enter the field of view. However, because both temporal motion estimation and disparity estimation are ill-defined problem, few prior researches have been devoted to the fusion of stereo disparity and 3D motion estimation.<sup>11,12</sup> In reality, the 3D facial model obtained from the first pair of image frames in a stereo image sequence is not accurate. The reasons are due to the possible occlusion and the inaccuracy in the feature point's position, disparity and camera model. To solve this problems, we must include more information, if any, from the other pair of stereo images.

#### 3.1. Off-Line Facial Model Estimation for Two Stereo Image Pairs

A facial model extraction (FME) algorithm for two pairs of stereo images is first proposed to solve these problems. Here we impose no constraint on the time distance between these two pairs of the stereo images and assume that the camera model is unknown. Since the time difference between the two pairs of stereo images may be large, the occlusion effects can be small and the model can be estimated more accurately. This proposed algorithm is not intended for real-time applications and thus named as off-line facial model estimation (off-line FME) algorithm.

##### Off-Line FME Algorithm

Assuming two pairs of stereo facial images are given. The images are  $I_{t,s}$ , where  $t = 1, 2$  denotes the first and second pair of stereo images and  $s = L, R$  denote the right or left view image frame. Thus we have four image frames,  $I_{1,L}, I_{1,R}, I_{2,L}, I_{2,R}$ . The relationship of disparity and 3D motion among these four image frames is illustrated in Fig. 3. In addition, we assume that the focal length of the lens in the imaging system and the distance and convergence angle between the two cameras are known. These are the parameters necessarily in a stereo imaging system. The following algorithm will estimate the facial model from these image frames if the perspective projection is used to model the image formation.

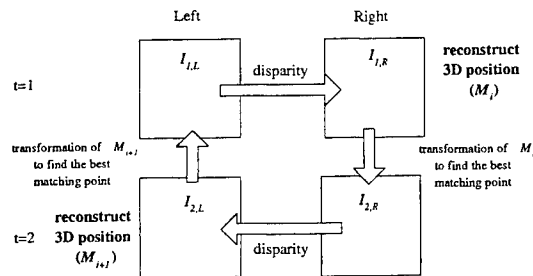
1. Locate the face and  $FP_{1,L}$  (the position of feature point  $F$ ) from image  $I_{1,L}$ . Here  $F$  can be any feature point shown in Figure 1. Use the feature point position  $FP_{1,L}$  as an initial guess to find  $FP_{1,R}$ ,  $FP_{2,L}$  and  $FP_{2,R}$  (the position of feature point  $F$  in image plane of  $I_{1,R}$ ,  $I_{2,L}$  and  $I_{2,R}$  respectively).
2. Estimate the facial model ( $M_1$ , the 3D positions of all the feature points) according to the feature points' position ( $FP_{1,L}$  and  $FP_{1,R}$ ), the disparity and focal length of the imaging system.
3. Find the rotation angle and translation distance of model  $M_1$  such that the 2D feature points' position in  $M_1$  and image  $I_{2,R}$  are best matched according to the perspective projection scheme.
4. Calculate the 2D perspective projective positions of all the feature points in  $M_1$  in the right image plane after rotation and translation and denote them as  $FP'_{2,R}$ .
5. Averaging  $FP'_{2,R}$  with  $FP_{2,R}$  to obtain the new estimated feature points  $\overline{FP}_{2,R}$ .
6. Estimate the facial model ( $M_2$ ) according to the feature points' position  $FP_{2,L}$ , estimated feature points ( $\overline{FP}_{2,R}$ ), disparity, and focal length of the imaging system.
7. Repeat the above procedures for image frames  $I_{1,L}$  and  $I_{1,R}$  to obtain facial model  $M_3$ .
8. Similarly, we have the facial model sequence  $M_i$ ,  $i = 1, 2, \dots$ . The computation of the facial model stops when  $M_i \approx M_{i+1}$ .

### 3.1.1. Transformation of Facial Model

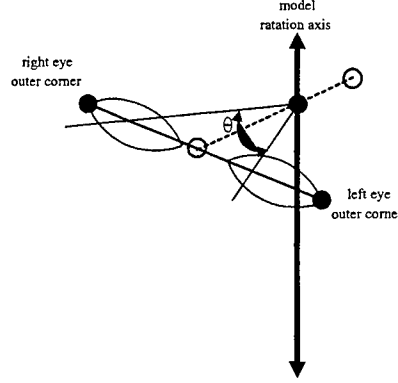
When the corresponding feature points of the stereo image pair,  $I_{t,L}$  and  $I_{t,R}$ , are detected, their 3D positions can be reconstructed. The rough model  $M_t$  can thus be established. So does another model  $M_{t+1}$  which are obtained from  $I_{t+1,L}$  and  $I_{t+1,R}$ . If the corresponding feature points are correctly extracted,  $M_t$  and  $M_{t+1}$  will be identical in this ideal case. Unfortunately, it is seldom to be so. The close-loop relationship of motion and disparity facilitates us to correct the errors derived from the mis-matched feature points. We will elaborate the this key technique of our model estimation algorithm in this section.

Analytically, the 3D motion of a rigid object can be decomposed as rotation and translation. Thus relation of  $M_t$  and  $M_{t+1}$  can be written as

$$M_{t+1} = \mathbf{R} \times M_t + \mathbf{T}, \quad (5)$$



**Figure 3.** Illustration of off-line FME algorithm through stereo image pairs



**Figure 4.** Transformation origin axis and rotation angle

where  $\mathbf{R}$  is the rotation matrix and  $\mathbf{T}$  is the translation matrix. Our purpose is to estimate the transformation,  $\mathbf{R}$  and  $\mathbf{T}$ , of the head in the time interval. Firstly,  $M_t$  should be translated and rotated, so that the perspective projections, denoted as  $FP'_{t+1,R}$ , is fitting the corresponding feature points in  $I_{t+1,R}$ , denoted as  $FP_{t+1,R}$ . The translation is regarded as the displacement between the centroid of  $FP'_{t+1,R}$  and the centroid of  $FP_{t+1,R}$ . It should be explained that if the translation in the Z coordinate exists, the head shown in the image frames of different time will have a scaling factor. But in our database, we impose no such condition. So the translation in the Z coordinate,  $T_Z$ , is set to be zero, and only the translations in the X and Y coordinate,  $T_X$  and  $T_Y$ , are calculated.

Choosing the correct location of rotation axis is required during rigid object transformation, but it is hard to make the decision without any priori knowledge about human motion mechanism. Thus a full search method is adopted to overcome the uncertainty. Here we set the rotation axis align with the middle point of the outer corner points of both eyes (feature points 3.7 and 3.12) in the vertical coordinate. But its depth is along the normal direction of the eye outer corners section toward the other side of the nose. The symmetry of human head is taken into consideration and the fact that the rotation axis locates in the inner skull is guaranteed. Figure 4 illustrate the relation of the eye corners and the rotation axis.

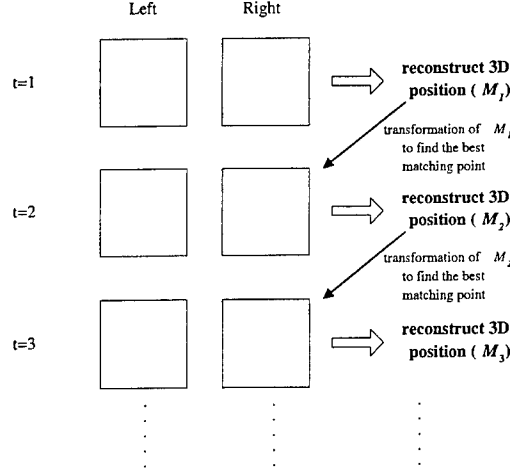
The variables in the full search method are the position of rotation axis and the rotation angle. The distance between the outer corners of both eyes is denoted as  $D_{corner}$ . Then the possible rotation axis position is located at the plane which is perpendicular to the outer eye corners connection and passes through the middle point,  $P_{mid}$ , of these two corners. Further, we consider the distance between the axis and  $P_{mid}$  is proportional to  $D_{corner}$ . We assume the range of the scaling factor is from 0 to 1, and the rotation angle is from  $-2/\pi$  to  $+2/\pi$ . The resolution is adjusted hierarchically to achieve a higher accuracy and better extremum. The initial search resolution is  $0.1 D_{corner}$  and  $1^\circ$ , and then increase the resolution to  $0.01 D_{corner}$  and  $0.1^\circ$  and so on. Finally, the distance between the projections of  $M_t$  ( $FP'_{t+1,R}$ ) and their corresponding points ( $FP_{t+1,R}$ ) is used as a decision measure.

### 3.1.2. Convergence of Iteration Process

The relative position error is used to check the convergence of model during the iteration process. Respectively, let  $F_{i,j}$  and  $C_i$  be the 3D position of the  $j$ th feature point and the centroid of all the feature point in model  $M_i$  at the  $i$ th iteration. The condition of convergence for our FME algorithm is

$$\frac{\sum_j [||F_{i,j} - C_i| - |F_{i+1,j} - C_i||]}{\sum_j |F_{i+1,j} - C_j|} \leq V_{th}, \quad (6)$$

where  $V_{th}$  is a pre-decided threshold value. The position error is the distance between the two corresponding feature points in the right and left view images. Due to the possible translation and rotation between the face in the first and second pair of stereo images, direct calculation of the position error suffers from the inherited bias term. To solve this problem, we thus make the centroid of all the feature points in the two models ( $M_i$  and  $M_{i+1}$ ) coincides and use relative error as single measure.



**Figure 5.** Illustration of on-line FME algorithm through a stereo image sequence

In summary, the proposed algorithm finds the ‘global motion’ of the head movement and the model parameters at the same time. Since the amount of global motion found (by any mean) will affect the precision of the model parameters, an algorithm is proposed to solve this problem iteratively. For two pairs of stereo images, the global motion between these two pairs is fixed and so is the model parameters. Therefore, the proposed algorithm will converge and the global motion and model parameters can be accurately estimated.

Throughout the development of this algorithm, the camera parameters we know are the focal length of the lens in the imaging system and the distance and convergence angle between the two cameras. The distance and convergence angle can be directly measured and the focal length can be obtained from the camera’s data sheet. Therefore no camera calibration is required but we can only obtain the ‘relative’ facial model where the unit of feature point’s position is pixel. If the relationship between the ‘unit’ in image plane and real world is known, then the exact facial model is obtained. To achieve this, camera calibration is necessary.

### 3.2. On-Line Facial Model Estimation for Stereo Image Sequence

Although the algorithm shown in the previous section is intended for off-line applications, it can be easily modified for on-line applications. That is, for stereo image sequence  $I_{t,i}$ , the off-line algorithm can be applied to images  $I_{t,i}$ ,  $t = 1, 2$  to obtain the estimated model. Subsequently, we have the second estimated model when  $t = 2, 3$  and so on. However, to do so requires a high speed computer because the algorithm requires several iterations to find the model. To solve this problem, we modify the off-line algorithm such that it can be used for on-line applications.

#### On-Line FME Algorithm

1. Same as the step 1 ~ 7 shown in the off-line FME algorithm for two stereo images.
2. Repeat the procedures for the image frames  $I_{3,L}$  and  $I_{3,R}$  to obtain facial model  $M_3$ .
3. Consecutively the following image pairs come in the processing. As a result, we have the facial model sequence  $M_i$ ,  $i = 1, 2, \dots$  in the forward direction. The computation of the facial model stops when  $M_i \approx M_{i+1}$ .

Above algorithm is illustrated in Fig. 5.

According to the  $FP_{1,L}$  and  $FP_{1,R}$ , the 3D position of facial model  $M_1$  is estimated. In the previous section, Equation (5) demonstrates that a deformation relation exists between  $M_1$  and  $M_2$ . We use a iterative full search method to explore the rotation angle and translation distance. This process is described in detail in the Subsection 3.1.3. Then the new estimated feature points  $\bar{FP}_{2,R}$  is calculated by averaging the projective points of  $M_1$  ( $FP'_{2,R}$ ) in the right image plane and the original feature points ( $FP_{2,R}$ ). A newly facial model  $M_2$  is thus estimated according

to  $FP_{2,L}$  and  $\overline{FP}_{2,R}$ . Repeating the procedure for  $M_2$  and the next image pair  $I_{3,L}$  and  $I_{3,R}$ , then estimating the facial model  $M_3$ . As the sequence goes on, we have the facial model sequence  $M_i$ ,  $i = 1, 2, \dots$ . If the convergence is achieved, which means that  $M_i \approx M_{i+1}$ , the computation of the facial model stops. The criterion of convergence, the same as one one in the off-line FME algorithm, is stated in Equation (6). Finally, a more accurate facial model  $\overline{M}$  is concluded by averaging the facial model sequence. But each model is obtained at different temporal point, there are slight difference between adjacent models. Directive averaging is not correct. We have to adjust these models so that they have the same direction. Since the rotation angle and translation distance are known in the above procedure, a inverse computation is applied to enforce every models face the same direction. Then the final estimated facial model  $\overline{M}$  is calculated as

$$\overline{M} = \frac{1}{N} \sum_i M_i. \quad (7)$$

#### 4. FACIAL MODEL ESTIMATION THROUGH MONO IMAGE SEQUENCE

In last section, we discuss the two proposed algorithms that can estimate a more accurate facial model using off-line and on-line techniques. But these two FME algorithms require stereo image sequence, the utilizations are limited due to the inconvenience of stereo image capture. In an ordinary case, only a single-view image sequence is obtained using the normal image capture systems, such as CCD camera or traditional VCR. This makes us to design a new facial model estimation algorithm, in which the only input is the mono image sequence. In a stereo image sequence, the redundancy comes from both disparity and motion, but in a mono image sequence, only the motion information is shown in the image content. Thereby a challenge is implied and deserves our efforts. Inheriting form the concepts in the FME algorithm through stereo image sequence, we tend to design the FME algorithm through mono image sequence in two aspects, off-line and on-line. The following sections give the details about the algorithms and their implementation procedures .

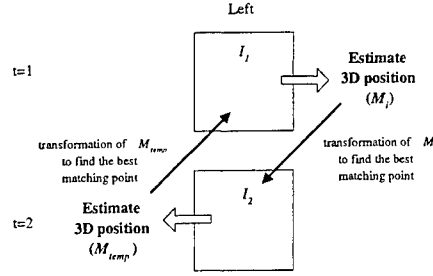
##### 4.1. Off-Line Facial Model Estimation for Two Mono Images

Here we impose no constraint on the time interval between these two images. Since the time interval between the two images may be large, the two images may have large discrepancy and thus the model can be estimated more accurately. This proposed algorithm is not intended for real-time applications and thus named as off-line facial model estimation (off-line FME) algorithm.

Assuming two facial images are given. The images are  $I_t$  where  $t = 1, 2$  denotes the first and second image. The following algorithm will estimate the facial model from these two image frames. Here orthogonal projection is used to model image formation because we assume the camera is directly in front of the face and the first image ( $I_1$ ) must contain the front view facial image. This implies the image plane (u,v) coincides with the  $X - Y$  plane of the model coordinate. The depth information in the  $Z$  axis is what we want to estimate here.

##### Off-Line FME Algorithm

1. Locate the face and  $FP_1$  (the position of feature point  $F$ ) from image  $I_1$ . Here  $F$  can be any feature point shown in Figure 1. Use the feature point position  $FP_1$  as an initial guess to find  $FP_2$  (the position of feature point  $F$  in image plane of  $I_2$ ).
2. Estimate the facial model ( $M_1$ , the 3D positions of all the feature points) according to the feature point's position ( $FP_1$ ) and the anthropometric estimation scheme shown in.<sup>13</sup>
3. Find the rotation angle  $\theta$  and location of rotation axis  $l_{x,z}$  of model  $M_1$  (with respect to its center) such that the 2D feature point's position from  $M_1$  (after rotation) and image  $I_2$  are best matched.
4. Rotate model  $M_1$  by  $\theta$  and then calculate the 2D positions of all the feature points (through projection) that are denoted as  $FP'_2$ .  $\overline{FP}_2$  is defined as the average of  $FP_2$  and  $FP'_2$ .
5. Combine the  $\overline{FP}_2$  and the depth of each feature point of the rotated model  $M_1$  to construct  $M_{temp}$ , the 3D position at second image time instance.
6. Repeat Step 3 such that the 2D feature point's position from ( $M_{temp}$ ) (after projection) and  $FP_1$  are best matched. Then use the idea in Steps 4 and 5 to find  $M_2$ .



**Figure 6.** Illustration of off-line FME algorithm through a mono image sequence

7. Keep looping the above procedures for images at  $t = 1$  and  $t = 2$ , we have facial model sequence  $M_i, i = 1, 2, \dots$ . The computation of the facial model stops when  $M_i \approx M_{i+1}$ .

Figure 6 shows the idea of this algorithm.

3D motion of a rigid body consists of rotation and translation. The relationship between the facial model  $M_i$  and  $M_{temp}$  is

$$M_{temp} = \mathbf{R} \times M_i + \mathbf{T}, \quad (8)$$

where  $\mathbf{R}$  and  $\mathbf{T}$  are the rotation and translation matrix, respectively. According to the relationship between model  $M_i$  and  $M_{temp}$ , we have

$$x_{temp} = x_i \cos \theta + z_i \sin \theta + T_x \quad (9)$$

$$y_{temp} = y_i \quad (10)$$

$$z_{temp} = -x_i \sin \theta + z_i \cos \theta + T_z \quad (11)$$

where  $(x_i, y_i, z_i)$  is the 3D position of the feature point  $F$ . Because the rotation axis is parallel to the  $Y$  axis,  $y_i$  and  $y_{temp}$  should keep the same. Since the orthogonal projection is employed in monoscopic case, we have the 2D position of the feature point  $F$  in  $I_1$  and  $I_2$  as  $(x_1, y_1)$  and  $(x_1 \cos \theta + z_1 \sin \theta, y_1)$ , respectively, where no head translation exists between these two time instances.

Since  $I_1$  contains the front view facial image, we can use the scheme shown in<sup>13</sup> to estimate the depth information  $z_1$ . Once the  $z_1$ , location of rotation axis  $l_{x,z}$ , rotation angle  $\theta$ , and the 2D position of feature point  $F$  at time  $t = 1, 2$  are known, we can update the position of feature point (in  $x$  and  $z$  coordinate) according to Equations 9-11. By the iteration process, an almost accurate position of the feature point  $F$  can thus be obtained.

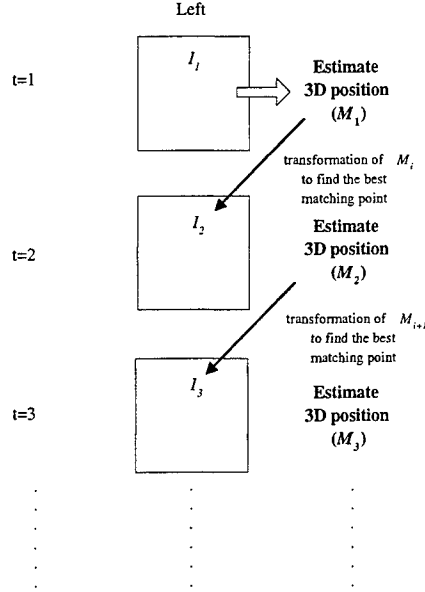
#### 4.2. On-Line Facial Model Estimation for Mono Image Sequence

If the input image sequence is plenty enough, the on-line FME algorithm is easily obtained after applying slight modifications from the off-line FME algorithm. The only difference between these two algorithms is at the step when the facial model of the second image is estimated. Off-line algorithm returns backward to the first image, but on-line algorithm seeks forward to the next image. The same problem comes from the different facing direction in the sequence content is addressed out and has to be solved.

##### On-Line FME Algorithm

Although the algorithm shown above is intended for off-line applications, it can be easily modified for on-line applications. That is, it is first applied to images  $I_1$  and  $I_2$ . Then images  $I_2$  and  $I_3$ , and so on. However, to do so needs a high speed computer because of the amount of computations required. To solve this problem, we modify the original algorithm for on-line processing. Considering the mono image sequence  $I_t$  where  $t = 1, 2, \dots$ , the on-line FME algorithm is shown below.





**Figure 7.** Illustration of on-line FME algorithm through a mono image sequence

1. Same as the step 1 ~ 4 shown in the off-line FME algorithm for two mono images.
2. Combine the  $\overline{FP}_2$  and the depth of each feature point of the rotated model  $M_1$  to construct  $M_2$ , the 3D position at second image time instance.
3. Repeat the above procedure to update model  $M_i$  from image  $I_i$ ,  $i = 3, 4, \dots$ . The computation of the facial model stops when  $M_i \approx M_{i+1}$ .

Figure 7 shows the idea of this algorithm.

## 5. RESULTS

In all the four algorithms proposed in this paper, we can only obtain the ‘relative’ facial model where the unit of feature point’s position is ‘pixel’ but not ‘mm.’ Therefore, a scaling factor must be found beforehand for accuracy evaluation. In this section the estimated value from our FME algorithm is attached with a hat, while the actual with no accessory, such as  $\hat{F}_j$  and  $F_j$  respectively.

Since the rotation axis, parallel to  $Y$  axis, was found in model estimation, we first define the shortest distance between the  $j$ th feature point ( $F_j$ ) and the rotation axis  $l_{x,z}$  as  $d_{xz}(F_j)$ . Then the scaling factor ( $SF$ ) of the feature point with respect to the rotation axis of the head is

$$SF_{xz} \equiv \frac{1}{N} \sum_j \frac{d_{xz}(F_j)}{d_{xz}(\hat{F}_j)} \quad (12)$$

where the unit of this scaling factor is  $mm/pixel$ .

Similarly, we define the vertical distance between a feature point ( $F$ ) and the topmost feature point 3.13 (or 3.14) in the facial model as  $d_y(F)$ . The scaling factor of the feature point with respect to the feature point 3.13 (or 3.14) is

$$SF_y \equiv \frac{1}{N} \sum_j \frac{d_y(F_j)}{d_y(\hat{F}_j)} \quad (13)$$

The overall scaling factor between the actual and estimated facial model is thus defined as the average of  $SF_{xz}$  and  $SF_y$ . Therefore we can scale the estimated model such that it is about the same size as the actual model. Then the estimated model after scaling is denoted as  $\hat{F}_j'$ , and its unit is 'mm.'

To know the accuracy of the model estimated, single measure must be defined. Respectively, let  $F_j$  and  $C$  be the actual 3D position of the  $j$ th feature point and rotation center and  $\hat{F}_j'$  and  $\hat{C}'$  denote the estimated 3D position (in mm). The relative error in the estimated facial model is then defined as

$$Error \equiv \frac{1}{N} \sum_j \frac{[|F_j - C| - |\hat{F}_j' - \hat{C}'|]}{|F_j - C|}. \quad (14)$$

Table 1 shows the facial model obtained from both off-line and on-line FME algorithms for stereo image sequence. The results convince us that the correct calibration of camera parameters is not important and an added advantage for the use of two pairs stereo images is to lessen the necessity of camera calibration. In addition, the redundant information in the image sequence helps to generate a improved facial model with less error.

The facial model derived from the mono image sequence is shown in Table 2, where the relative positions of all the feature points are listed. Comparing with the results shown before, the on-line algorithm still finds the accurate position of all the feature points (within 6.56% error). In addition, for mono images, on-line algorithm provides results with smaller error compared with the off-line algorithm because of the additional information available in the image sequence.

According to the accurate positions of all necessary feature points, a 3D mesh model is thus generated. Then the first image  $I_1$ , which is the front-view facial image under the beginning assumption, is applied as a texture material in the texture mapping procedure. The mapping method is the UV plane mapping. Figure 8 shows the estimated facial model (after texture mapping) from two mono image images. Although this is the largest error feature point set, the results resemble true face very well. And the other three algorithms show the better performance.

## 6. CONCLUSION

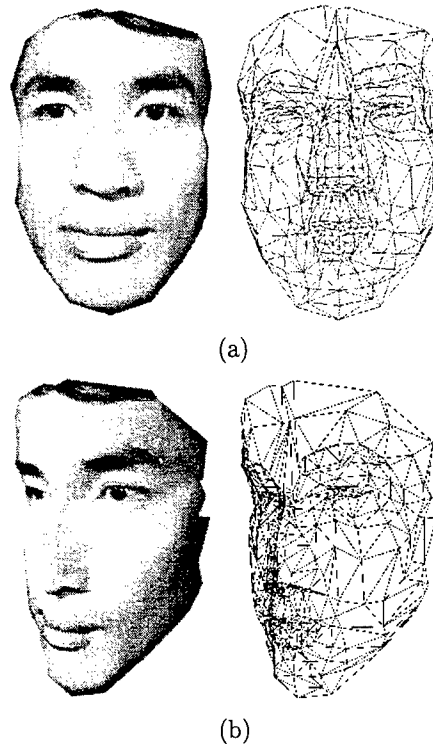
Extraction of facial model from stereo image sequence is important for MPEG-4 related applications. Conventional approaches estimate the facial model from only one pair of stereo images but suffer from the occlusion effects and tedious camera calibration. In this paper, we propose the off- and on-line facial model estimation algorithms to estimate the facial model from stereo image sequence. Simulation results show that our schemes can obtain a much more accurate facial model compared with the conventional approach without the camera calibration.

Our schemes are suitable for practical on- and off-line applications. For off- line applications, two pairs of stereo facial images are first captured and our algorithm can be used to obtain the facial model without camera calibration. With slight modifications of the off-line FME algorithm, our algorithm can also be used for on-line facial model estimation.

Moreover, we extend the application to the mono image sequence. It was shown in<sup>13</sup> that facial model can be estimated from single front-view facial image. However, some model parameters (that is, the positions of some feature points) suffer large estimation error. Our subsequent works concentrate on the modification of the proposed algorithm here for mono image sequence. Inherited from the concepts of on- and off-line in the stereo case, both algorithms are established. The accuracy of the proposed FME algorithms through mono image sequence depends a lot on the initial facial model estimated from the first image. As a consequence the first image comes into the mono FME algorithms has to be chosen carefully. The lower error results convince us that the proposed facial model estimation algorithms through a mono image sequence are practicable.

## REFERENCES

1. R. Schäfer & T. Sikora, "Digital video coding standards and their role in video communication," *Proceedings of the IEEE*, vol. 83, pp. 907-924, June 1995
2. "MPEG-4 Overview (Tokyo Version)," *Coding of moving pictures and audio*, ISO/IEC JTC1/SC29/ WG11 N2196, Mar. 1998



**Figure 8.** The estimated facial model from two mono images. (a)frontal view, and (b)45-degree view

3. S.C. Chang, K. Aizawa, H. Harashima & T. Takebe, "Analysis and synthesis of facial image sequences in model-based image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, pp. 257-275, Jun. 1994
4. G. Galicia & A. Zakhor, "Depth based recovery of human facial features from video sequences," *Proceedings of IEEE International Conference on Image Processing*, pp. 603-606, Oct. 1995
5. R.C. Gonzalez & R.E. Woods, *Digital Image Processing*, Reading: Addison- Wesley, 1992
6. K. Aizawa, T. Saito & H. Harashima, "Model-based analysis synthesis image coding (MBSAIC) system for a person's face," *Journal of Signal Processing: Image Communication*, vol. 1, pp. 139-152, Oct. 1989
7. H. Tao, and T. S. Huang, "Deriving Facial Articulation Models from Image Sequences," *Proceedings of IEEE International Conference on Image Processing*, October 1998.
8. I. A. Kakadiaris and D. Metaxas, "Model-Based Estimation of 3D Human Motion with Occlusion Based on Active Multi-Viewpoint Selection," *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 81-87, June 1996
9. J.M. Chung & T. Nagata, "Binocular vision planning with anthropomorphic features for grasping parts by robots," *Robotica*, vol. 14, pp. 269- 279, 1996
10. T. Nguyen, T. Huang, "Segmentation, grouping and feature detection for face image analysis," *Proceedings of IEEE International Symposium on Computer Vision*, pp. 593-598, Nov. 1995
11. X. Chen & A. Luthra, "MPEG-2 multi-view profile and its application in 3DTV," *SPIE Proceedings*, vol. 3021, pp. 212-223, 1997
12. W. Richard, "Structrue form stereo and motion," *Journal of the Optical Society of America A*, vol. 2, pp. 343-349, Feb. 1985
13. C.J. Kuo & R.S. Huang, "Synthesizing lateral face from frontal facial image using human anthropometric estimation," *Proceedings of IEEE International Conference on Image Processing*, vol. 1, pp. 133-136, Oct. 1997

**Table 1.** Feature point's positions estimated from the proposed FME algorithm through stereo image sequence.

Feature Point	Off-Line FME Algorithm	On-Line FME Algorithm
3.12	(-94.073 -60.022 195.38)	( -96.87 -60.037 197.03)
3.14	(-62.504 -73.052 199.14)	( -63.97 -72.85 203.57)
3.8	(-35.005 -58.284 196.24)	(-36.368 -58.489 201.34)
3.10	(-63.651 -54.208 202.46)	(-63.957 -54.392 205.52)
3.6	(-63.075 -66.904 200.79)	(-63.942 -66.707 207.19)
3.11	( 40.235 -62.093 196.31)	( 40.939 -61.914 199.18)
3.13	( 67.457 -74.789 194.22)	( 68.802 -74.207 198.5)
3.7	( 99.858 -59.153 192.48)	( 99.498 -58.875 195.71)
3.9	( 67.302 -57.015 197.75)	( 69.555 -56.583 201.26)
3.5	( 66.311 -67.907 197.55)	( 68.569 -67.53 201.03)
8.4	(-51.181 78.165 200.12)	(-51.139 78.37 204.4)
8.9	(-20.651 59.522 234.6)	(-18.039 59.318 232.74)
8.1	(-6.1504 64.2 238.66)	(-2.4026 64.206 233.77)
8.10	( 7.4639 59.322 239.15)	( 11.021 59.298 233.41)
8.3	( 54.356 79.168 210.72)	( 55.481 79.122 211.34)
8.2	(-0.1476 108.77 227.39)	( 2.3189 108.98 229.43)
2.2/2.3	(-4.4316 79.969 236.75)	(-1.2229 80.01 232.95)
9.4	(-42.211 27.648 207.93)	(-41.266 27.386 209.53)
9.3	(-16.832 16.756 269.69)	(-9.9246 16.032 255.48)
9.5	( 41.061 23.104 221.62)	( 42.896 22.203 222.25)
9.15	(-8.1349 36.803 244.43)	(-4.6879 36.657 238.8)
Error	4.6011%	3.9459%

**Table 2.** Feature point's positions estimated from the proposed FME algorithm through mono image sequence.

Feature Point	Off-Line FME Algorithm	On-Line FME Algorithm
3.12	(-93.92,-60.1,187.2)	(-95.6,-59.38,184.2)
3.14	(-63.63,-72.4,231.9)	(-66.14,-71.62,231.1)
3.8	(-34.62,-58.03,163.6)	(-36.76,-57.34,166.9)
3.10	(-62.63,-54.04,215.1)	(-64.92,-53.26,218.4)
3.6	(-61.13,-67.16,189.6)	(-63.15,-66.52,198)
3.11	(38.35,-62.02,164.4)	(35.48,-61.42,167.4)
3.13	(63.01,-74.26,191.1)	(59.93,-73.66,198.2)
3.7	(93.39,-57.97,187.4)	(90.03,-57.67,187)
3.9	(62.38,-56.81,229.2)	(58.78,-56.32,234.1)
3.5	(60.79,-67.16,240.5)	(57.06,-66.52,244.1)
8.4	(-50.27,77.77,196.3)	(-52.71,76.98,199.9)
8.9	(-10.71,59.11,234.9)	(-14.05,58.62,237.1)
8.1	(6.159,64.33,203.5)	(3.212,63.72,209.5)
8.10	(18.27,59.24,226.6)	(14.87,58.62,228.1)
8.3	(54.82,78.74,207.4)	(51.6,78,209.9)
8.2	(6.097,108.6,254.8)	(2.072,107.6,255)
2.2/2.3	(5.409,79.75,240)	(1.975,79.02,241.9)
9.4	(-40.82,27.22,234.8)	(-43.85,26.99,234.1)
9.3	(2.333,16.67,285.4)	(-1.855,16.45,282)
9.5	(45.08,21.95,225.4)	(41.49,21.55,227.9)
9.15	(4.903,36.52,229.2)	(1.624,36.17,232.2)
Error	7.21%	6.56%

# Vision-Based Intelligent Robots

Minh-Chinh Nguyen

Institute of Measurement Science, LRT6  
Federal Armed Forces University Munich  
85577 Neubiberg - Germany  
E-mail: Minh.Chinh@Unibw-muenchen.de

## ABSTRACT

Vision is an ideal sensor modality for intelligent robots. It provides rich information on the environment as required for recognizing objects and understanding situation in real time. Moreover, vision-guided robots may be intelligent and largely calibration-free, which is a great practical advantage. Together with it, a new concept for intelligent robot control, that enables realization of the calibration-free visual robots, is introduced.

**Keywords:** Calibration-Free Robots, Vision-Guided Intelligent Robots, Robot Vision, Situation-Oriented and Behavior-Based Robot Control.

## 1. INTRODUCTION

Industrial robots are of great economic and technological importance. Until 1996 approximately 860,000 robots had been installed worldwide. At that time 680,000 of them were still being used, for the most part in automobile and metal-manufacturing [IFR, 1997]. Typical applications include welding card, spraying paint on appliances, assembling printed circuit boards, loading and unloading machines and placing cartons on a pallet. Experts estimate that by the year 2000 about 950,000 industrial robots will be employed world-wide.

Although present robots contribute very much to the prosperity of the industrialized countries they are quite different from the robots that researchers have in mind when they talk about "intelligent robots". Today's robots are not creative or innovative, do not think independently, do not make complicated decision, do not learn from mistakes and do not adapt quickly to changes in their surroundings. They rely on detailed teaching and programming and carefully prepared environments. It is costly to maintain them and it is difficult to adapt their programming to slightly hanged environmental conditions or modified tasks.

Although the vast majority of robots today are used in factories, advances in technology are enabling robots to automate many tasks in non-manufacturing industries such as agriculture, construction, health care, retailing and other services. These so-called "field and service robots" aim at the fast growing service sector and promise to be a key product for the next decades.

From a technical point of view service robots are intermediate steps towards a much higher goal: "personal robots" that will be as indispensable and ubiquitous as personal computers today. Personal robots must operate in varying and unstructured environments without needing maintenance or programming. They must cooperate and coexist with humans who are not trained to cooperate with robots and who are not necessarily interested in them. Advanced safety concepts will be as indispensable as intelligent communication abilities, learning capabilities, and reliability. It will be a long way of research to achieve this goal, but undoubtedly vision - the most powerful sensor modality known - will enable these robots to perceive their environments, to understand complex situation and to behave intelligently.

This paper present some of the underlying concept and principle that were key to the design of our research robots. It is organized as follows: in the next chapter will be briefly described the vision and its potential for robots. The third chapter will describe the new concept for intelligent robot control. The experiments and results as well as conclusions will be discussed in the fourth and fifth chapter respectively.

## 2. VISION AND ITS POTENTIAL FOR ROBOTS

### 2. 1. Advantages of the visual sensors and a conceptual structure of robot's vision systems

When a human drives a vehicle he depends mostly on his eyes for perceiving the environment. He uses his sense of vision not only for locating the path to be traversed and for judging its condition, but also for detecting and classifying external object, such as other vehicles or obstacles, and for estimating their state of motion. Entire situations may thus be recognized, and expectations, as to their further development in the "foreseeable" future, may be formed.

The same is true for almost all animals. With the exception of those species adapted to living in very dark environments, they use vision as the main sensing modality for controlling their motions. Observing animals, for instance, when they are pursuing prey or trying to escape a predator, may give an impression of the performance of organic vision system for motion control.

In some modern factory and office buildings mobile robots are operating, but almost all of them are blind. Their sensors are far from adequate for supplying all the information necessary for understanding a situation. Some of them have only magnetic or simple optical sensors, allowing them merely to follow an appropriately marked track. They will fail whenever they encounter an obstacle and they are typically unable to recover from a condition of having lost their track. The lack of adequate sensory information is an important cause making these robots move in a comparatively clumsy way and restricting their operation to the simplest of situations.

Other mobile robots are equipped with sonar systems. Sonar can, in principle, be a basis for powerful sensing systems, as evidenced by certain animals, such as bats or dolphins. But the sonar systems used for mobile robots are usually rather simple ones, their simplicity and low cost being the very reason for choosing sonar as a sensing modality. It is then not surprising that such system are severely limited in their performance by low resolution, specular reflections, insufficient dynamic range, and other effects.

Nevertheless, even when comparing the most highly developed organic sonar systems with organic vision systems, it is obvious that in all environments where vision is physically possible animals endowed with a sense of vision have, in the course of evolution, prevailed over those that depend on sonar. This may be taken as an indication that vision has, in principle, a greater potential for sensing the environment than sonar. Likewise, it may be expected that advanced robots of the future will also rely primarily on vision for perceiving their environment, unless they are intended to operate in other environments, e.g. under water, where vision is not feasible.

One apparent difficulty in implementing vision as a sensor modality for robots is the huge amount of data generated by a video camera: about 10 million pixels per second, depending on the video system used. Nevertheless, it has been shown (e.g., by [Graefe 1989]) that modest computational resources are sufficient for realizing real-time vision systems if a suitable system architecture is implemented.

As a key idea for the design of efficient robot vision systems the concept of object-oriented vision was proposed. It is based on the observation that both the knowledge representation and the data fusion processes in a vision system may be structured according to the visible and relevant external objects in the environment of the robot (Figure 1). For each object that is relevant for the operation of the robot at a particular moment the system has one separate "object process". An object process receives image data from the video section (camera, digitizers, video bus etc.) and generates and updates continuously a description of its assigned physical object. This description emerges from a hierarchically structured data fusion process which begins with the extraction of elementary features, such as edges, corners and textures, from the relevant image parts and ends with matching a 2-D model to the group of features, thus identifying the object.

This concept is practical because it was found that in any given moment only a small number of objects are relevant and that, consequently, only a small number of processes need to be active simultaneously. In the next moment, however, different objects may be relevant; therefore, the ability to switch the system's focus of attention quickly is crucial. The switching of attention and the control of the cameras is performed by a vision system management process that dynamically generates appropriate object processes upon request.

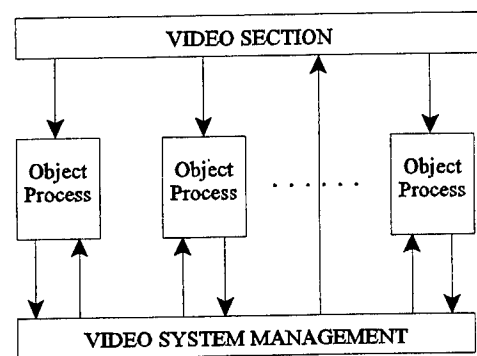


Figure 1: Conceptual structure of object-oriented robot vision system

The potential of object-oriented vision systems was first demonstrated in high-speed autonomous highway driving applications [Graefe, Kuhnert 1988], [Graefe 1992]. Later the same concept has proved its value in mobile and stationary indoor robots.

## 2. 2. Perception

Model-based robot control depends on a continuous flow of numerical values describing the current state of the robot and its environment. These values are derived from measurements performed by the robot's sensors. One problem here is that the quantities that are needed for updating the numerical models may be difficult to measure, e.g., the distance, mass and velocity of some external object that is posing a collision danger. Also, there are certain important decisions that cannot be made on the basis of measurements alone; the hypothetical decision whether in a particular situation a collision with a parked car should be brought about in order to avoid a collision with a pedestrian is an example.

Humans and other organisms, on the other hand, do not depend on measurements for controlling their motions. If, for instance, we want to sit down on a chair or pass through an open door, we do not first measure the size of the chair, the door, or our body; rather, we make a qualitative judgement whether the chair is high or low, or whether the door is wide or narrow, and then execute a sequence of motions that is adequate for the situation. In short, we substitute perception for measurement.

According to Webster's Dictionary "perception" is:

- ▶ a result of perceiving;
- ▶ reaction to sensory stimulus;
- ▶ direct or intuitive recognition;
- ▶ the integration of sensory impressions of events in the external world by a conscious organism;
- ▶ awareness of the elements of the environment.

"To perceive" means, according to the same source,

- ▶ to become aware of something through the senses;
- ▶ to become conscious of something;
- ▶ to create a mental image;
- ▶ to recognize or identify something, especially as a basis for, or as recognized by, action.

Typical questions to be answered by perception are:

- ▶ Which objects exists?
- ▶ What is the relationship between objects?
- ▶ Is it necessary to react? How?

Perception, rather than measurement, is thus a prerequisite for, and a complement of, situation assessment. Vision is the ideal sensing modality for perception because it is capable of supplying very rich information on the environment.

The actual design and implementation of a behavior pattern and of related perceptual processes depend on the robot's environment and task. A mobile robot navigating in a network of passageways needs different behaviors and recognition modules than a walking robot intended to explore rough terrain.

However, advantage sensor system will be got their efficiency fully, if and only if they are combined with a sensible control concept. In the sequel we will, so that, represent a new concept of "behavior-based and situation-oriented robot control" for intelligent vision-guided robot control.

### 3. CONCEPT OF SITUATION-ORIENTED AND BEHAVIOR-BASED VISUAL ROBOT CONTROL

#### 3.1. Behavior

Biological behaviors could be defined as any thing that an organism does involving action and response to stimulation, or as the response of an individual, group, or species to its environment. Behavior-based robotics has become a very popular field in robotics research because biology proves that even the simplest creatures are capable of intelligent behavior: They survive in the real world and compete or cooperate successfully with other beings. Why should it not be possible to endow robots with such an intelligence? By studying animals behavior, particularly their underlying neuroscientific, psychological and ethological concepts, robotic researchers have been enabled to build intelligent behavior-based robots according to the following principles:

- ▶ complex behaviors are combinations of simple ones, complex actions emerge from interacting with the real world
- ▶ behaviors are selected by arbitration or fusion mechanisms from a repertoire of (competing) behaviors
- ▶ behaviors should be tuned to fit the requirements of a particular environment and task
- ▶ perception should be actively controlled according to the actual situation

Many system architecture and control methods, which were introduced in recent years, interest in realizing of behavior-based robots. Its main characteristics are active perception of the robot's dynamically changing environment, recognition and evaluation of its current situation, and dynamic selection of behaviors appropriate for the actual situation. Animals simplest capabilities, i.e., to perceive and act within an environment in a meaningful and purposive manner, can thus be imitated by our robots to a certain degree.

#### 3. 2. Situation Assessment

According to the classical approach, robot control is model-based. Numerical models of the kinematics and dynamics of the robot and of the external object that the robot should interact with, as well as quantitative sensor models, are the basis for controlling the robot's motions. The main advantage of model-based control is that it lends itself to the application of classical control theory and, thus, may be considered a straight-forward approach. The weak point of the approach is that it breaks down when there is no accurate quantitative agreement between reality and the models. Differences between models and reality may come about easily; an error in one of the many coefficients that are part of the numerical models suffices. Among the many possible causes for discrepancies are initial calibration errors, aging of components, changes of environmental conditions, such as temperature, humidity, electromagnetic fields or illumination, maintenance work and replacement of components, to mention only a few. Consequently, most robots work only in carefully controlled environments and need frequent recalibrations, in addition to a cumbersome and expensive initial calibration.

Organisms, on the other hand, are robust and adapt easily to changes of their own conditions and of the environment. They never need any calibration, and they normally do not know the values of any parameters related to the characteristics of their "sensors" or "actuators". Obviously, they do not suffer from the shortcomings of models-based control which leads us to the assumption that they use something other than numerical models for controlling their motions. Perhaps their motion control is based on a holistic assessment of situation and the selection of behaviors to be executed on that basis, and perhaps robotics could benefit from following a similar approach.

According to Webster's Third New International Dictionary [Babcock 1976] the term "situation" describes among others "the way in which something is placed in relation to its surroundings", a "state", a "relative position or combination of circumstances at a given moment" or "the sum of total internal and external stimuli that act upon an organism within a given time interval". We define the term "situation" in a similar way, but with a more operational aim, as the set of all decisive factors that should ideally be considered by the robot in selecting the correct behavior pattern at given moment. These decisive factors are:

- ▶ perceivable objects in the environment of the robot and their suspected or recognized states;
- ▶ the state of the robot (state of motion, presently executed behavior pattern, ...);
- ▶ the goals of the robot, i.e., permanent goals (survival, obstacle avoidance) and transient goals emerging from the actual mission description (destination, corridor to be used, ...);



- ▶ the static characteristics of the environment, even if they cannot be perceived by the robot's sensors at the given moment;
- ▶ the repertoire of available behaviors and knowledge of the robot's abilities to change the present situation in a desired way by executing appropriate behavior patterns.

Figure 2 illustrates the definition of the term "situation" by embedding it in the action- perception loop of a behavior-based and situation-oriented robot. The actions of the robot change the state of the environment, and some of these changes are perceived by the robot's sensors. After assessing the situation an appropriate behavior is selected and executed, thus closing the loop. The role of a human operator is to define external goals via a man machine interface and to control behavior selection, e. g., during supervised learning.

Although situation-oriented robot control has proven much more robust and flexible under real-world conditions than classical model-based control it is not perfect. One reason is that, obviously, the robot cannot base its behavior selection on a "true" or "real" situation, but only on an internal image of the situation as created by the robot according to its sensor information and its - always imperfect - knowledge of the world and of its own characteristics. Also, disturbances during the behavior execution can lead to non-expected situations. Although the disturbances may be corrected by either adjusting behavior-immanent parameters or selecting a different behavior, they will usually cause the robot to move in a non-ideal way.

#### 4. IMPLEMENTATION

The described concepts were implemented on the calibration-free vision-guide manipulator Mitsubishi Movemaster RV-M2 with 5 degree of freedom (Figure 3) for grasping objects of various shapes (Figure 4). It eliminates the need for a calibration of the robot and of the vision system, it uses no world coordinates, and it comprises an automatic adaptation to changing parameters. The concept is based on the utilization of laws of projective geometry that always apply, regardless of camera characteristics, and on machine learning for the acquisition of knowledge regarding system parameters. Different forms of learning and knowledge representation have been studied, allowing either the rapid adaptation to changes of the system parameters or the gradual improvement of skills by an accumulation of learned knowledge.

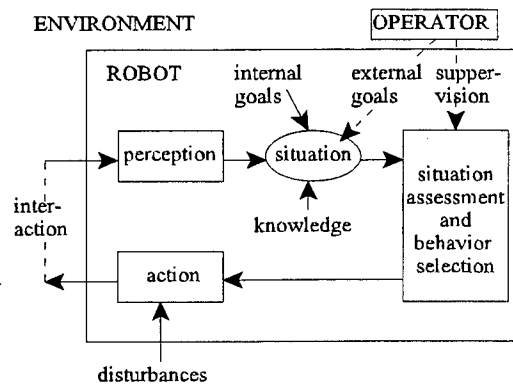
The images from the two cameras are processed by an object-oriented vision system described in 2.1 above, which consists of two frame grabbers, each containing a TMS320C40 Digital Signal Processor.

The situation process receives and assesses the information about the position and orientation of gripper and of object to be grasped to decide which behaviors of the robot [Nguyen 1999] will be used to achieve the grasp, and to generate appropriate motion control commands.

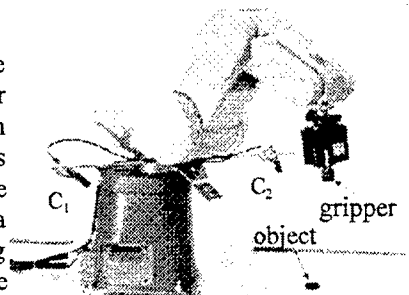
#### 5. EXPERIMENTS AND RESULTS

The described concepts has been evaluated in a series of real-world experiments.

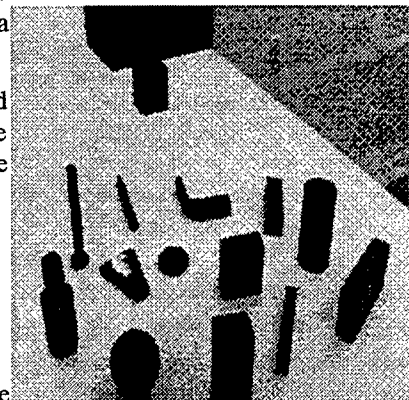
Objects of various shapes were successfully grasped. It requires no knowledge regarding:



**Figure 2 :**  
The role of "situation" as a key concept in the perception action loop of a situation- oriented behavior-based robot.



**Figure 3:** External view of the Movemaster RV-M2 with mounted cameras.



**Figure 4:** Objects used in our experiments

- ▶ The parameters of the robot arm
- ▶ The internal camera parameters, i.e., optical characteristics
- ▶ The exact locations of the cameras  
(except that the cameras should be located some distance away from the work plane of the robot in an opposite arrangement)
- ▶ The exact viewing directions of the cameras  
(except that both cameras should have the actual work space of the robot in their fields of view)
- ▶ The dimensions, kinematics, and joint angles of the robot  
(except that, for practical reasons, we presently assume that the robot is of an articulated arm type, and that the general type of the gripper and the number of degrees of freedom of the system are known)
- ▶ The quantitative relationships between the control words sent to the motor controllers and the resulting motions  
(except that these relationships are assumed to be "smooth")
- ▶ The surrounding environment, e.g., lighting, surrounding landmarks  
(except that it should be within reason)

In addition, even severe disturbances, such as arbitrary changes of the cameras' orientations, that would make other robots fail, are tolerated while our robot is operating.

We state that the concepts proposed in this work will be especially valuable for mobile and service robots operating in unstructured environments.

## 6. CONCLUSIONS

A fundamental concepts and principles for realization of intelligent robots have been presented. We strongly believe that vision - the sensor modality that predominates in nature - is also an eminently useful and practical sensor modality for intelligent robots. It provides rich and timely information on the environment and allows real-time recognition of dynamically changing situations. Situation-dependent perception and behavior selection rather than measurement and control based on quantitatively correct models are additional key factors for advanced robots. Motor control commands should be derived directly from sensor data, without using world coordinates or parameter-dependent computations, such as inverse perspective or kinematic transforms.

Building robots according to these rules and testing them intensively in the real world lead to robust and intelligent robots with the ability to adapt themselves to modified environmental conditions and tasks.

## REFERENCES

1. P. Babcock, *Webster's Third New International Dictionary of the English language*, G. & C. Merriam Company, Springfield, MA, USA, 1976.
2. J. R. Cooperstock, E. E. Milios, "Self-supervised learning for docking and target reaching," *Robotics and Autonomous Systems 11*, pp 243-260, 1993.
3. V. Graefe, K.-D. Kuhnert, "Towards a Vision-based Robot with a Driver's License," *Proc. IEEE Int. Workshop on Intelligent Robots and System, IRO '88*. Tokyo, pp. 627 - 632, 1988.
4. V. Graefe, "Dynamic Vision Systems for Autonomous Mobile Robots," *Proc. IEEE/RSJ International Workshop on Intelligent Robots and Systems, IROS '89*, Tsukuba, pp. 12-23, 1989.
5. V. Graefe, "Visual Recognition of Traffic Situations by a Robot Car Driver," *Proceedings, 25<sup>th</sup> ISATA ; Conference on Mechatronics*, Florence, pp 439 - 446, 1992.
6. V. Graefe, Q.-H. Ta, "An Approach to self-learning Manipulator Control Based on Vision," *Proc. International Symposium on Measurement and Control in Robotics*, pp 409-414, Smolenice, 1995.

7. IFR International Federation of Robotics "Key data for the world robot market," [Http://www.ifr.org/stat.htm](http://www.ifr.org/stat.htm), 1997.
8. I. Kamon, T. Flash, S. Edelman, "Learning to Grasp Using Visual Information," *Proc. IEEE International conference on Robotics and Automation*, vol. 2, pp 2470-2476, 1996.
9. M.-C. Nguyen, V. Graefe, "Object Manipulation Controlled by Uncalibrated Stereo Vision". *The Second Chinese Congress on Intelligent Control and Intelligent Automation; Proceeding of the CWCICIA '97*, Vol. 1, Xian-China, pp. 77-83, 1997.
10. M.-C. Nguyen, "Situation-oriented and Behavior-based Stereo Vision to Gain Robustness and Adaptation in Manipulator Control," *In D. Casasent (ed.): Intelligent Robots and Computer Vision XVIII: Algorithms, Techniques, and Active Vision, Proceedings of the SPIE*, Vol. 3837, pp. 90-97, Boston, USA, 1999.
11. K. Vollmann, M.-C. Nguyen, "Manipulator Control by Calibration-Free Stereo Vision". *In D. Casasent (ed.): Intelligent Robots and Computer Vision XV, Proceedings of the SPIE*, Vol. 2904. Boston-USA, pp. 218-226, 1996.

# The Face Recognition Security Entrance

C. W. Ni<sup>1</sup>

Opto-Electronics & Systems Laboratories  
Chutung, Taiwan, Republic of China

## ABSTRACT

This paper describes an automatic face recognition algorithm for security entrances. There are two major steps in this procedure to make the automatic recognition possible:

1. We combined the two-phase face detection method and back propagation neural networks to detect human faces when people are walking in the region of entrances. The combination allows the strength of both methods activated to accommodate the size and head-orientation variations and to eliminate the false detection.
2. Novel face recognition: we extract the facial feature measurements to form the multi-variable normal distribution for each person. These multi-variable normal distributions separate the decision space well and the probability for good index for face recognition.

This face recognition algorithm is very efficient on computing time and taking little storage space.

**Keywords:** Face Recognition, Feature Extraction, Back Propagation Neural Networks, Multi-variable Normal Distribution

## 1. INTRODUCTION

We design a security entrance to allow authorized personnel un-interruptedly walking into the entrance. The user's face image was taken during the early stage in the one's entering. The one's facial information is analyzed and identified from the authorized personnel database. Currently, there exist security systems designed with the combination of the password and the face recognition. But only a face recognition technique, which may work in clustered moving backgrounds and can recognize the face from a database, is able to provide user-friendly applications.

The field of face detection and face recognition have increasingly attracted researchers' interest over the past a few years. While the face detection is the foundation of face recognition. For human intelligence, the face detection is a easy task; but for the artificial intelligence, it takes quiet a while to do the computing to detect a face. There are several face detection and face recognition image processing techniques under development<sup>1</sup>: neural network based technique<sup>2-4</sup>, feature-based technique<sup>5-6</sup>, the view-based eigenspaces technique<sup>7-9</sup>, the elastic matching technique<sup>10,12</sup>, three-dimensional surface-based approaches<sup>11</sup>, and face recognition using Hidden Markov Models<sup>13</sup>.

The Neural network-based algorithms<sup>2-4</sup> reported fair results. Lin *et al.* used a Probabilistic Decision-Based Neural Network approach reported fast and nearly 100% recognition rate. Lawrence *et al.*, combined a self-organizing map neural network and a convolutional neural networks. Rowley *et al* first did the histogram equalization to correct lighting variations, and employed various window sizes and different subsampling ratio to detect all possible faces, 90.5% recognition rate was achieved.

The feature-based<sup>5-6</sup> face recognition by fitting deformable templates to face regions to extract the face feature geometry, then to normalize them for standardization. Normalized features are classified by a set of principle eigenvectors. The representation vectors of this classification can be mapped to the facial features independent of the facial expressions.

---

<sup>1</sup> Further author information -  
Email: cwni@oes.itri.org.tw

Shackleton *et al* used six potential energy terms to include all variables, in an iteration process of templates fitting, iteratively modifying all parameters. But this method didn't reach very high recognition rate. Cox *et al*, therefore, adopted a mixture-distance techniques, each face represented by 30 distances calculated from 35 measured positions, reached 95% recognition rate.

The view-based eigenspaces<sup>7-9</sup> assumes that the set of all possible face patterns occupies a small and parameterized subspace. The approach approximates the subspace of face patterns using data clusters and principal components from one or more example sets of face images. An advantage of eigenspace approaches is that it is possible to deal with occluded situations. The approach is only demonstrated to be working in un-clustered background. With the likelihood estimate which can be made optimal (with respect to information-theoretic divergence) and can be computed solely from the low-dimension subspace projection coefficients, thus yielding a computationally efficient estimator for high density probability density function. The eigenfaces face detection technique is a fast, simple, and practical method. But the detection highly depends on the high correlation between the training images and the test images.

The elastic matching technique, employs the Gabor decomposition, accommodates a shift, dilation or local transformations (such as head tilting or smiling)<sup>11</sup>. The Gabor decomposition of an object image  $I(x)$ , obtained by convolving it with the complex Gabor filter kernels  $G_{\Psi(k,l)}$  is an iconic multiresolution template. To reduce the interpixel redundancy, subsampling this template forms a Gabor grid  $\mathcal{G} = \{V', E'\}$  that covers the object with  $N \times M$  nodes (vertices  $V'$ ) in the  $x$  and  $y$  directions, and edges ( $E'$ ), respectively. The magnitude of the Gabor probe is used to measure the similarity between matched local features, while the phase of Gabor probe is used to fine tune the matching result. The extracted information (both signal energy and local pattern structure) associated with each probe spans a multiresolution neighborhood whose size equals the extent of the filter kernels. Each Gabor probe records Gabor decomposition of an object at a spatial location  $x$  with spectral extent  $\Delta\omega_k$ . It also represents the fact that the low-frequency channel extracts coarse image features in a large neighborhood, while high-frequency channel can extract fine localized features in a small neighborhood. The grid node records the localized feature at each spatial location, and the grid-edge records the spatial relationship between nodes<sup>11</sup>.

The three-dimensional surface-based approach model the frontal (physical) human faces as a 3D surface and recover it from its 2D image with 3D eigen head of eigensurfaces values which obtained from the laser scans of human faces. The eigen head method can exclude the lighting condition variations, but it cannot not deal effectively with rotation<sup>10</sup>. The Hidden Markov Model, with the combination with neural networks, applied on face recognition<sup>13</sup> reported 100% recognition rate. The newly developed algorithms have the character that combine several methods such that each method can take advantage of its strength, and the slightly overlapped calculation insures better algorithm robustness.

## 2. THE FACE RECOGNITION ALGORITHM

By analyzing the image taken from the CCD and matching with the data base of the authorized personnel, the user's access of the security entrance is authorized (Figure 1). Since the analyzing images are taken in a region in front of the entrance, the sizes of the face images have both upper and lower bound. We adapt the Two-Phase face detection algorithm to catch a suitable size of a face in a clustered background. This detection algorithm tolerates some face size and orientation variations, but not quite enough for the image of a walking user stepping into our entrance.

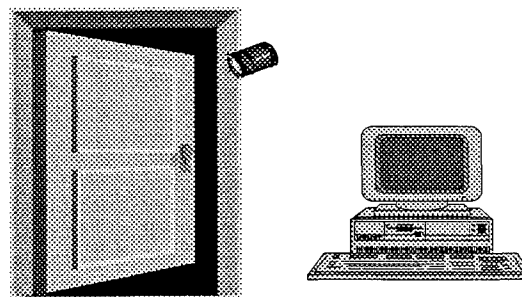


Fig. 1

In order to catch all possible size of face images in the defined region, for a fixed size of template, the images are subsampled into different resolutions. To accommodate user head's tilting, in the image taken processes, the images are rotated in each admissible angle. The two-phase face detection algorithm<sup>1</sup> is employed for each size and each angle of images (Figure 2).



Fig. 2

Then we send the set of detected faces into a pre-trained back propagation neural networks to find the best faces in terms of face sizes and face orientations. In Figure 2, the best face is the one in second row, first column. The neural network, used to train the system, is a three layer backward propagation neural network. There are 12, 4, 1 cells in each of the layer, respectively. The logsig function is used as the transfer function. The network was first trained by a batch wok, then fine tuned by incremental training with other variations, such as glasses wearing faces, big round-eye faces, long faces, round faces etc.

We recognize the best-detected faces to decide whether the accesses are granted. The number of sizes is determined by the image-taken-region in front of the entrances. The number of different face orientations is determined by the degree of the up-right of the users' faces we expect. This detection algorithm makes no pre-defined-assumptions on the background. For example, first, we have a face image with moving clustered background (as in Fig. 3):



Fig.3

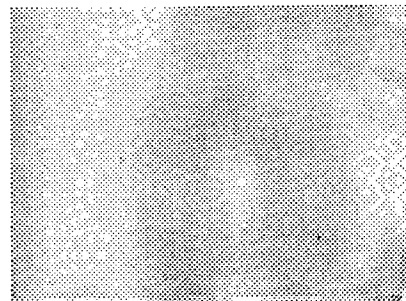


Fig. 4

Then, we want to find a face which includes from the eyebrow to the upper lips, the correlation coefficient result calculated in Figure 4, as in two-phase face detection algorithm<sup>1</sup>, the binarized matching results are shown in Figure 5, and the calculated possible facial region is shown in Figure 6.

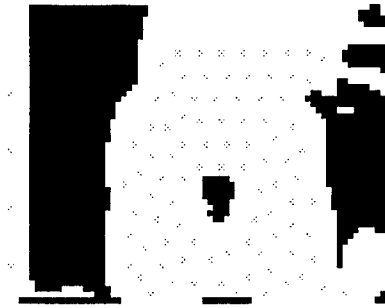


Fig. 5

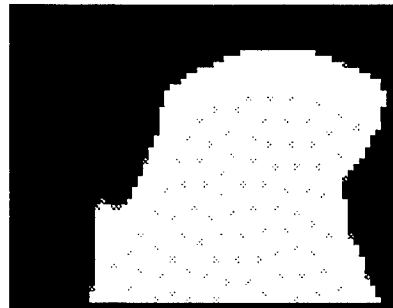


Fig. 6

To combine Figure 4 and Figure 6, the “face” is detected in Figure 7. Have the detected face in full resolution, Figure 8, the eyes and the eyeballs and the nose are located and length measured.

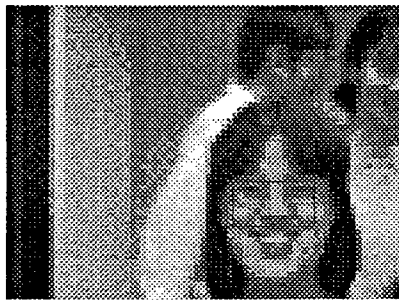


Fig. 7

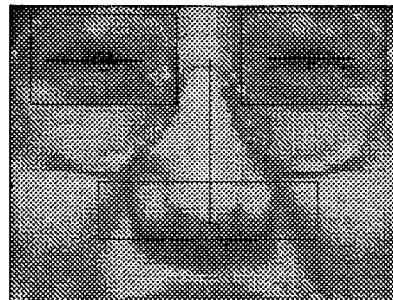


Fig. 8

For the techniques mentioned in the first section, the recognition can be decomposed as representation and matching. A good representation of an object carries the enough information toward the goal, but also saves the storage space and the computing time. While a good matching algorithm takes the minimum information produces a well-separated decision space for decision-making.

Since human being carries different habits and tendency. The variations of their features are quite different too. To collect all authorized personnel information in one space, the information of each person's each feature variation can be imbedded in the variation of multivariable-normal distributions. The metric of the decision space, the distance of the sample point to each person may be defined by each person's multivariable probabilities.

A simple feature as facial length measurements is employed to recognize faces for a small number of authorized personnel. The length measurements of each person is roughly normal distribution: Figure 9 shows the distance between two eyes (from the left eye inner point to the right eye inner point) are fairly good Gaussian distribution. The Figure 10 and Figure 11 shown left and right eye length, respectively, the normality test is good too.

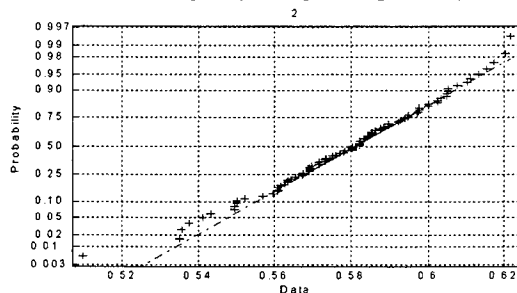


Fig. 9

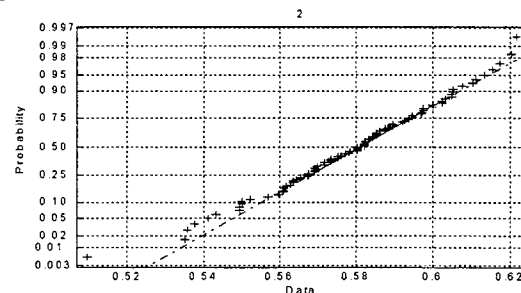


Fig. 10

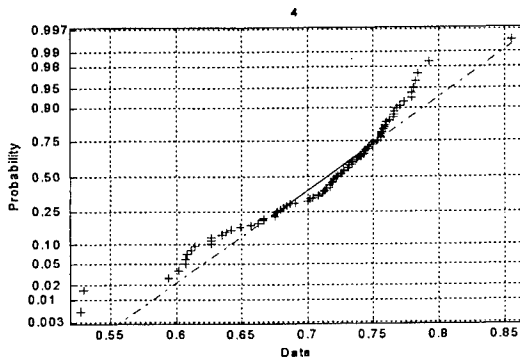


Fig.11

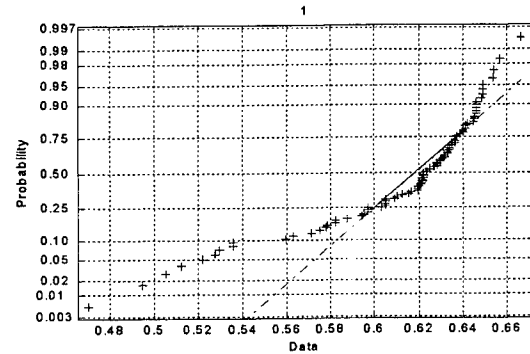


Fig. 12

Figure 12 is the eyeball-distance normality test. Figure 13 is the nose length normality test. While the Figure 14 is the normality test for nose width. The nose width and the eyeball distance are not quite well fit into normal distribution, especially, the nose width. Since the Nose is a 3D feature in the face, the face forward direction has non-linear influence on the nose width measure.

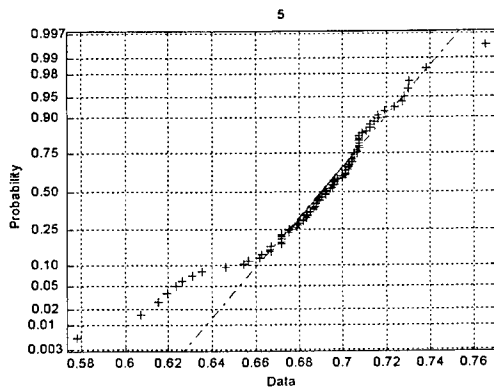


Fig. 13

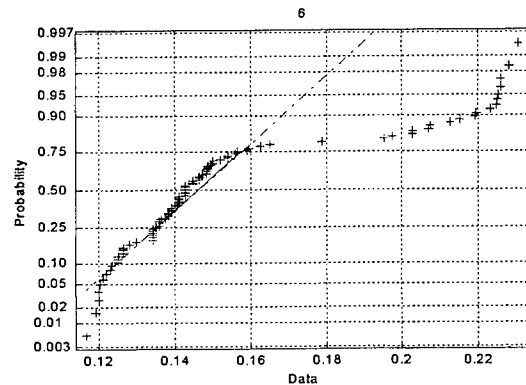


Fig. 14

Theoretically speaking, we can increase the number of features to increase the recognition population, therefore, split the nose width into 3 measures may improve the fitting. Then with the set of features, the multi-variable normal distribution of each person is determined by analyzing a sample series of facial image.

### 3. RESULTS

The program constitutes two parts (Fig. 15): building database for authorized personnel and face recognition. We take ten length measurements in each face. In the “building database for authorized personnel” part, we take 35 “good” (some face image are not quite frontal views which cause the measurements abnormal) samples for each person to calculate the multi-variable normal distribution as personal parameters. When the person is not facing the camera or the face is not detected, the algorithm will continuously take images and do facial measurements analysis, then recognize the face. In each “face recognition”, the facial measurements taken from the face image are used to calculate the personal



probability for every person in the database. When the maximum conditional probability is greater than the threshold value, the face image is identified.

With Matrox Meteor II frame grabber on Pentium II, clock speed 400 MH, 192 MB RAM, the visual C++ program did the face recognition in 3 to 4 seconds in each round. We project this algorithm ought to be fast enough for real time operation with DSP parallel processing chip. We obtained 86% correct recognition rate among 5 persons, each person with 30 times of recognition. For there are still room for the improvements of the facial length measurements taken and measurement choices, we expect the algorithm is able to do better.

#### 4. SUMMARY

In this paper, we proposed a face recognition algorithm for the security entrance systems. This algorithm does not interrupt the proceeding of the users, does not need the users pay special attention to the security system, as long as the users don't turn their faces away from the Image taking camera during the image-taken region. There is no assumption on the background. This algorithm also tolerates the up-right variation of the face-tilting as the face-tilting angle desired. This algorithm will suffice the small-authorized personnel entrances, such as small companies, laboratories, or home, small dormitories.

There are a few things can improve the result further. As the Figure 14 indicates the nose width is a poor length measurement choice for normal distribution. While nose width is an important feature in human faces. Use EM method to model the human nose width may be better improve the recognition results.

#### 5. ACKNOWLEDGEMENTS

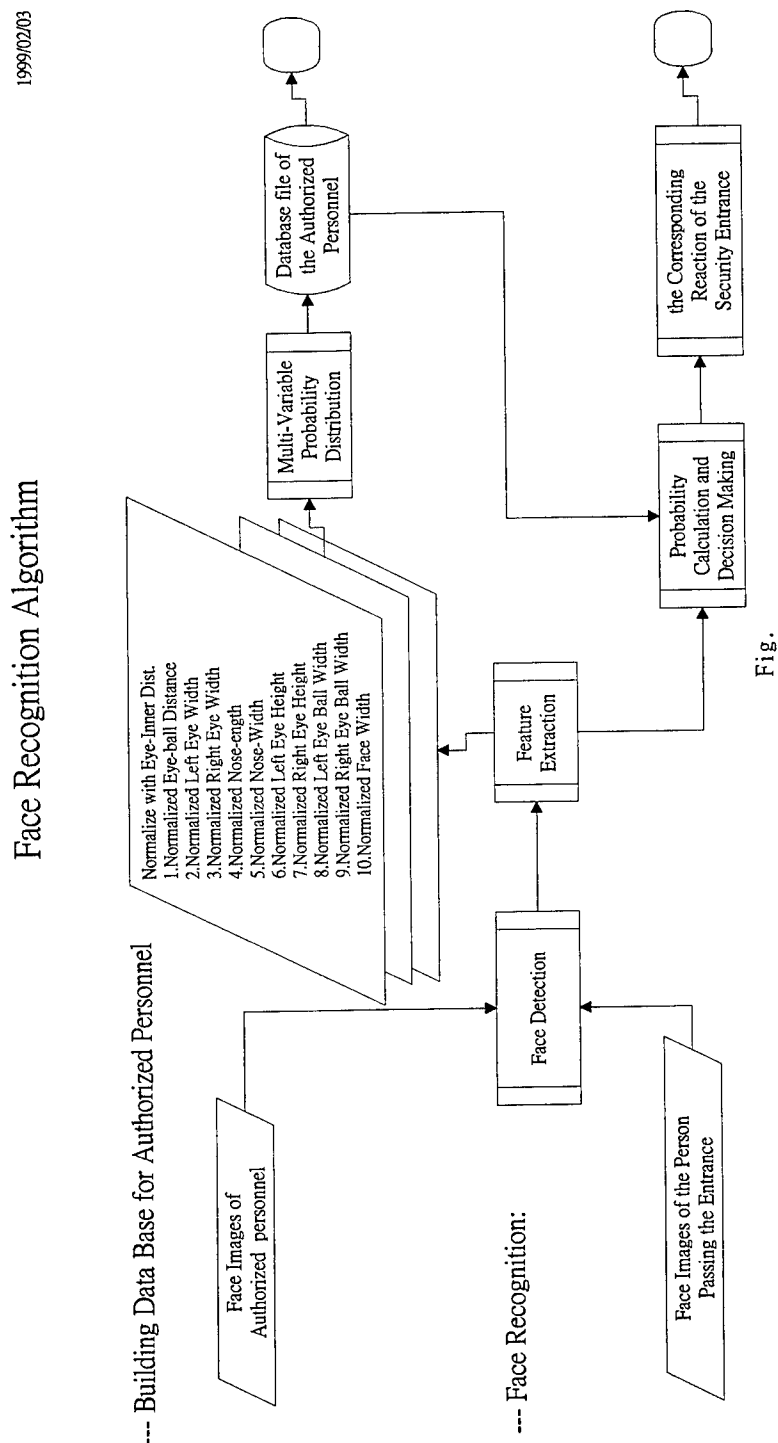
This work was supported by the Opto-Electronics & Systems Laboratories, the Industrial Technology Research Institute, Republic of China.

#### 6. REFERENCES

1. C. W. Ni, "The Tracking 3D Display Systems by Image Processing Analysis", *SPIE Proceedings* vol. 3296, pp. 217-224, Jan. 1998.
2. Shang-Hung Lin, Sun-Yuan Kung, and Long-Ji Lin, "Face Recognition/Detection by Probabilistic Decision-Based Neural Network," *IEEE trans. Neural Networks*, Vol. 8, pp. 114-132, Jan. 1997.
3. S. Lawrence, L. Giles, and A. C. Tsoi, "Convolutional Neural-Network for Face Recognition," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 217-222, 1996.
4. H. A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1996.
5. I. J. Cox, J. Ghosn, and P. N. Yianilos, "Feature-Based Face Recognition Using Mixture-Distance", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1996.
6. M. A. Shackleton, and W. J. Welsh, "Classification of Facial Features for Recognition", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1991.
7. M. A. Turk, and A. P. Pentland, "Face Recognition Using Eigenfaces", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1991.
8. T. Darrell, B. Moghaddam, and A. P. Pentland, "Active Face Tracking and Pose Estimation in an Interactive Room", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1996.
9. M. A. Turk, and A. P. Pentland, "Face Recognition System", *U. S. Patent, No. US5164992*, 1992.
10. Jun Zhang, Yong Yan, and M. Lades, "Face Recognition: Eigenfaces, Elastic Matching, and Neural Nets," *Proc. Of the IEEE*, Vol. 58, No.9, pp. 1423-1435, Sept.1997.
11. L. P. Ammann, "Robust Image Processing for Remote Sensing Data", *Proceedings of 1<sup>st</sup> International Conference on*

Image Processing, Austin , pp. 41-45, 1994.

12. Xing Wu and Bir Bhanu, "Gabor Wavelet Representation for 3-D Object Recognition", *IEEE trans. on Image Processing*, Vol. 6, pp. 47-64, Jan. 1997.
13. K. S. Yoon, Y. K. Ham, and R.-H. Park, "Hybrid Approaches to Frontal View Face Recognition Using the Hidden Markov Model and Neural Networks", *Pattern Recognition*, Vol. 31, no. 3, pp. 283-293, 1998.



# **Seagle-1, A new man-portable thermal imager**

Ruey-Nan Yeh, Fu-Fa Lu, Hong-Ming Hong, Ya-Tung Cherng and Horng Chang

Materials & Electro-optics Division

Chung Shan Institute of Science and Technology

P.O.Box90008-8-7, Lung Tan, Taoyuan Taiwan 325

## **ABSTRACT**

This paper presents the design and performance of Chung Shan Institute of Science and Technology (CSIST) newest man-portable infrared imaging system, the **Seagle-1**.

The thermal imager is designed for day and night long-range observation and forward reconnaissance and surveillance applications. The camera system achieves an NETD = 0.067K at 30 Hz frame rate with f/1.8 optics (300K background). The design and performance of the 256x244 PtSi infrared camera will be described in this paper.

**KEYWORDS:** platinum silicide, infrared , stirling mincooler

## **1. INTRODUCTION**

The thermal imager system incorporates a number of advanced features to achieve lower power, compact size, and high performance. The sensor assembly is a 256x244 MWIR Platinum Silicide schottky barrier staring FPA, integrated to a low power high reliability stirling minicooler , some of the system features include : lightweight, f/1.8 100mm telescope internal 2.5" LCD min-monitor or external video output RS-170, and low power electronics. The nonuniformity calibration values are programmed at the factory and require no additional recalculation in the field. This feature offers significant saving in space and power. The system operates from standard internal battery or external power. Operation time by using the internal rechargeable Ni-MH battery is 3 hours. This system is quiet, rugged, reliable, easily maintainable, and affordable. Because of the compact size, low power consumption, and high performance, Seagle-1 is well suitable for portable applications.

## **2. CAMERA SYSTEM DESIGN**

The unique design feature of the Seagle-1 was described in this section.. The design feature was based upon customer demanding a thermal imaging system with high sensitivity in a single compact package. The package outline was shown in Figure 1. The electrical power dissipation within the system is nominally 26 Watts when operating at 23°C ambient. The ruggedized housing was constructed by using aluminum alloy. Integral hard mount surfaces for the printed circuit boards and cryocooler/detector assembly provide effective heat sinking via conduction. No force convection cooling is required allowing the unit to be completely sealed and resistant to rain and moisture. All mechanical parts, including the lens assembly, were "O-ring" sealed and maintained at a positive pressure inside the case. The camera include the lens weight was

less than 3.0kg. The imager includes the detector/cooler assembly 、 electronics subassemblies 、 F.L100 mm f/1.8 standard lens 、 2.5" LCD display assembly and housing components. This is significant because nonuniformity calibration values that are programmed at the factory and require no additional recalculation in the field,.This feature offers significant compact in space and power saving.



Figure 1. SEAGLE-1 camera system

## 2.1 System Features

System features are summarized as follows

Sensor Material	PtSi Schottky-Barrier IRCCD
Image Format	256Vx244H
Spectral range	3.4~5 $\mu$ m
Optics	4.6° X3.5° with 100mm lens, operation.....manual
Cold Shield	f 1.8 100% Efficient
Cooling method	Stirling-cycle cooler
Uniformity Calibration	2 point
Video format	RS-170, B&W
Weight	3.0 Kg
Power Supply	Rechargeable Battery or External AC
Data Output	12bit digital data,External Video connector

## 2.2 256x244 PtSi CCD Imager

The 256x244 element PtSi SB IRCCD detector is configured with monolithic focal plane array. The dimension of each pixel is 31.5x25 $\mu$ m<sup>2</sup> and a fill factor greater than 36%. The double poly single metal process and 1.5 $\mu$ m design rule was used on these Pt/Si chips. The chip is monolithic silicon design using an interline transfer charge coupled device (CCD) readout

architecture. The readout is carried out by vertical and horizontal CCD which is driven by four-phase clock pulses. The output preamplifier is a floating diffusion amplifier (FDA) with a two-stage source follower. The output transfer conversion gain of the FDA was greater than  $2.0\mu\text{V}/\text{electron}$ . The sensor was backside illuminated and sensitive to radiation in the  $1\text{--}5\mu\text{m}$  wavelength band. The specifications of the focal plane array and measured performance were summarized in Table 1.

Figure 2. was the PtSi chip configuration. The chip size and active image area were  $9.4 \times 7.7 \text{ mm}^2$  and  $8.1 \times 6.1 \text{ mm}^2$  respectively. The device is bonded in a 32-pin ceramic package with a hole in the center for backside illumination.

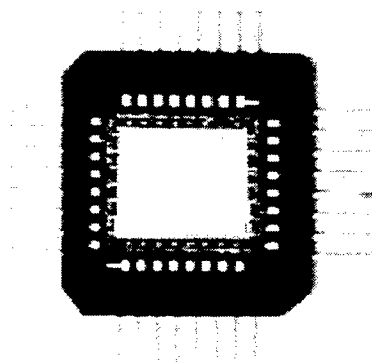


Figure 2. Photograph of 256x244 PtSi IRCCD sensor

Table 1. Focal plane Array Specification

Material	Platinum Silicide
Array Size	256 H x 244 V
Detector size	$19.5 \mu\text{m} \times 17 \mu\text{m}$
Pixel size	$31.5 \mu\text{m} \times 25 \mu\text{m}$
Fill Factor	>36%
Infrared Band	MWIR
Architecture	Charge Coupled Device
VCCD & HCCD	4 Phase
Charge Capacity	$1 \times 10^6$ electrons per pixel
Charge Transfer Efficiency	>0.9998
Quantum Efficiency	>0.6 % at $4 \mu\text{m}$

### 2.3 Optics

The optic system consist of a four-element , refractive ,100mm focal length, MWIR lenses, in conjunction with a silicon dewar window,  $3.4\mu\text{m}$  to  $5\mu\text{m}$  cold bandpass filter, and a F/1.8 cold shield.

### 2.4 Detector- Dewar Cooler Assembly

The crycooler-dewar assembly was a permanently evacuated dewar with an internal f/1.8 cold shield and cold filter cooled by a miniature integral 0.5 W Stirling cooler. In order to minimize size and power dissipation within the sensor assembly, an

Integrated Detector Cooler Assembly (IDCA) is utilized. In an IDCA packaging scheme, the focal plane array, cold shield and cold filter were attached directly to the crycooler expander. The IDCA packaging technique eliminates the significant thermal losses and thermal mass of a dewar inner stem. At the system level, the result is lower input power and faster cool down time for a given cooler. The miniature integral 0.5 W Stirling cycle cooler was the Model K-508A which manufactured by Ricor Ltd. The K508 model was the newest microcooler product offered by Ricor. The integration of the detector/dewar assembly was shown in Figure 3. Although these coolers utilize a rotary type compressor, it demonstrated a reliability of greater than 4000 hours. These coolers was suitable applied for portable/low power systems. The power efficiency and audible noise requirements were obviously extremely critical to hand held, battery operated imaging system, particularly for both military and law enforcement application.

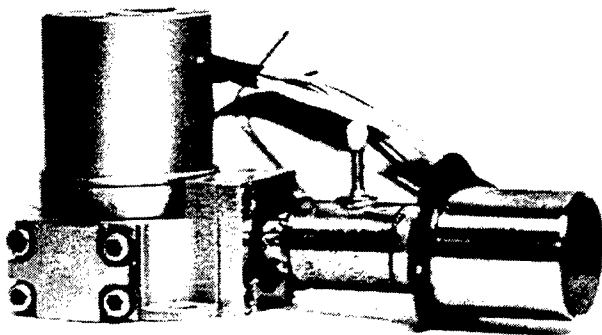


Figure 3. Integral Dewar/Cooler Assembly

## 2.5 Electronics

The electronics were design to accommodate the weight, power and functionality goals of Seagle-1. A generalized schematic of the electronics signal flow was shown in Figure 4. The major functions were implemented in two primary circuit boards. The first primary circuit board was the drive and power supply board. It contains circuitry to operate the FPA and generate NTSC video timing and signal levels. This drive board also apply contrast and brightness adjustments. The signal was digitized into 12 bit digital video signals which sent to digital signal processing (DSP) board. The power supply circuit received a 28 DCV input and produce all the necessary voltages for the electronics. It also supplies power to the 2.5" LCD display monitor.

The second primary circuit board was the data board. It contains circuitry for nonuniformity correction. The correction method used here is a two-point correction scheme. The gain and offset for each pixel are measured in lab and stored in EPROM. This approach required no additional recalculation in field operation, and offered significant saving in space and power. After the correction, the processed digital video is sent back to digitizer board again and was converted to standard RS-170 TV compatible signal.



### 3.1 NETD

The NETD of those camera have measured from the signal swing level to a temperature difference ( $\Delta T = \pm 5^\circ\text{C}$ ) and noise level. The following is the measurement result of NETD and condition of measurement. The measured NETD of Seaeagle-1 at 25 C was 0.067 C with a F/1.8, 100 mm lens.

### 3.2 MRTD

The MRTD is defined as the minimum temperature difference of the 4-bar target, which is resolvable at each spatial frequency. We measured the horizontal and vertical MRTD in standard condition (with 100mm lens and at 23C background temperature). The horizontal MRTD of the Seaeagle-1 is 0.278 C and the vertical MRTD is 0.299 C both at 1.66 cycles/mRad (Nyquist frequency of Seaeagle-1). The horizontal MRTD of SeaEagle-1 is 0.092 C and the vertical MRTD is 0.105 C both at 0.804 cycles/mRad (about 1/2 Nyquist frequency). These results showing that the Seaeagle-1 has well consistent with original design. The MRTD test result was shown in Fig. 6.

### 3.3 Dynamic range (D/R)

The Dynamic range (D/R) is defined as the ratio of the temperature window to NETD. i.e.  $D/R = 20 \cdot \log(\text{Temp\_window}/\text{NETD})$ . The SeaEagle-1 thermal imager has a fixed gain for special usage. The temperature window is set to 16 Kelvin and D/R is 50 dB.

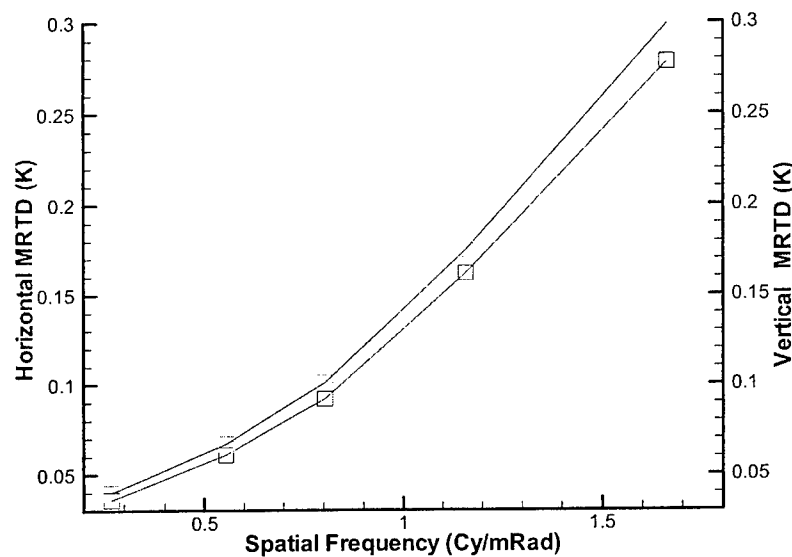


Fig.6. Seagle-1 horizontal and vertical MRTD



The summary of camera performance was shown in table 2

Table 2. Summary of the Seaeagle-1 thermal imager performances

Item	Performance
Model / Type	Sea Eagle-1
Detector	PtSi 256x244
NETD	0.067 C at 300K F/1.8
Horizontal MRTD	0.092 C at 0.5 Nyquist
Field of View	4.5° 100 mm lens
Analog video output	RS-170 (BNC) 3" LCD
Cooling method	Stirling-cycle cooler
Dimension (WxLxH)	108mm x 185mm x 116 mm without Lens
Weight	3 kg

#### 4. Conclusion:

The Seagle-1 infrared thermal image camera based on CSIST's PtSi SBFPA has been developed. The performance of the system was in consistent with the original design. This system was qualified and field tested. This compact system could be applied in surveillance and law enforcement.

#### REFERENCES

1. R.N.Yeh et al."High- performance 256x244 PtSi Schottky-Barrier IRCCD Imager" proceeding of the SPIE Vol.3377,P148-154,1998.
2. W.S.Wang, C.Ho, T-M.Chuang "High -performance IR detectors fabricated by PtSi on p-si substrate" proceeding of the SPIE Vol. 3379, P 333-343, 1998.
3. W. F. Kosonocky, "Review of Schottky Barrier Imager technology" proceeding, of the SPIE Vol.1308, P. 2-26, 1990.
4. D. L. Clark et al., "Design and Performance of a 486x640 Pixel Platinum Silicide IR Imageing system" proceeding of the SPIE Vol. 1540. P.303. 1991.
5. S. Fujino, T. Miyoshi, M.Yokoh and T.Kitahara,"Mitsubishi Thermal Imager Using the 512x512 PtSi Focal Plane Arrays" Proceeding of the SPIE Vol. 1157,1989.
6. N. B. Stetson, J. W. Landry "Handheld Imaging Using PtSi., InSb and HgCdTe Focal Plane Technology" SPIE July 1994.
7. Brian Toft et al., "Design of infrared camera utilizing a miniature crycooler with an integrated dewar/cooler assembly", Proceeding of SPIE Vol. 1675, 1992
8. J. W. Landry, N.B. Stetson., "Miniaturized Platinum Silicide Focal Plane Array Camera", SPIE Vol. 2225.
9. Y.J.Shaham, M. Umbricht, S. Rudin., "Cold Shield Effectiveness in MWIR Cameras" SPIE Vol. 2269 1994.
10. Masayuki Inoue et al., "Portable high performance camera with 801x512 PtSi-SB IRCSD" SPIE Vol.3061, P.150-158, 1997.



## **SESSION 2**

### **Output Devices and Imaging**

# New method of large ink supply without long tubing system for wide-format inkjet printer

Chin-Tai Chen\*

Printing Technology Division  
Opto-Electronics & System Labs  
Industrial Technology Research Institute  
Bldg. 78,K120 OES/ITRI, Chutung 310, Taiwan, R.O.C.

## ABSTRACT

This paper explored the design of large ink supply system of ink-jet printer, which is in general installed in so-called wide format printer today. Subsequently, a new type of large ink supply system (LISS) was presented to fulfill the fundamental functions that ink in the reservoir could be automatically delivered into the print head of printer by means of capillary force in nature. Moreover, the new system was characterized with no traditional long tubing portions such that pressure loss or vibration, due to long ink passage of tubing, could be eliminated. To achieve the goal of removing traditional long tubing system, the ink reservoir width of system must be greater than the print width of printer. As a result, a stable back pressure of print head can be kept all the time no matter the printer is printing or not; it could also be more stable than before even if print head is moving in high speed. Hence, better print quality could be obtained in the printer equipped with the new system of paper.

**Keywords:** Ink supply system, Ink-jet printer, Pressure, Ink reservoir

## 1. INTRODUCTION

In the past over 10 years<sup>1</sup>, many types of wide format printer (WFP) have already been successfully developed to print on different media with large width. Of course, both of sheet media and roll media can be alternatively used in the WFP. In general, the print width for those WFP could be 24 inches, 36 inches, 54 inches, or even possibly greater than 60 inches. It could easily found that most of WFP in common comprise of off-board ink supply system<sup>2,3,4,5,6,7,8</sup>. In addition, these ink supply system always use long and flexible tubes for delivering ink flow to the print head, so that two major problems may unfortunately exist thereof. One issue is pressure loss that may be caused by the movement of ink flow through the long tubing system; therefore, pressure wave could be induced as well at the same time. The loss might be expressed as equation (1) where  $L$  is the length of tube,  $V$  is the flow speed, and  $A$  is the cross sectional area of tube. It can be clearly understood from considering the effect of friction between the flow and tube. The other is pressure vibration that may mainly contribute from the movement of print head and tubing system. It can be simply described as equation (2) where  $L$  is the length of tube in moving and  $A$  is the cross sectional area of tube. Here, Newton's second law of force can explain well such an effect of unsteady pressure. In the mean time, it's noted as shown in Figure 1.0 that both of issues might be happening due to the long tubing system.

$$P_{loss} \propto \frac{L \times V}{A}, \quad (1)$$

$$\Delta P_{vibration} \propto L \times A, \quad (2)$$

---

\* Correspondence: E-mail: chintai@itri.org.tw; Telephone: 886-3-5918358; Fax: 886-3-5917446

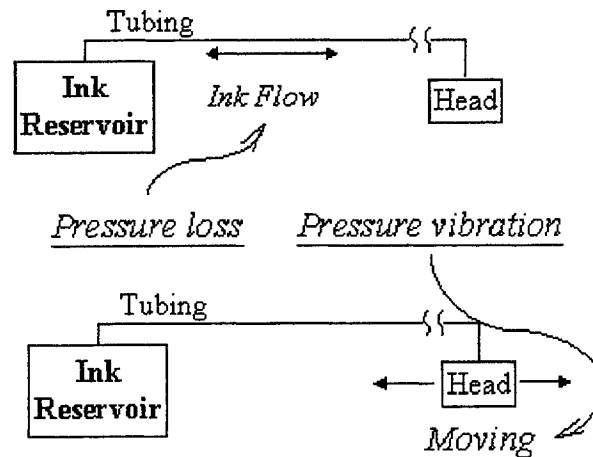


Fig. 1.0: Pressure loss and vibration due to the tubing system

General speaking, the pressure loss and vibration might be up to several centimeters of  $H_2O$  under the normal operation of WFP, respectively. Both of them are not desirable because print quality might be badly unstable. However, it's obvious that reducing the length of tubing or further even removing the long tubing system could eliminate both.

## 2. NEW DESIGN OF SYSTEM

One new design of system for ink supply was presented to solve the above-mentioned issues. The mechanical configuration of design was clearly demonstrated in the following Figure 2.0 where key components of WFP were already set up, together with ink supply system thereof.

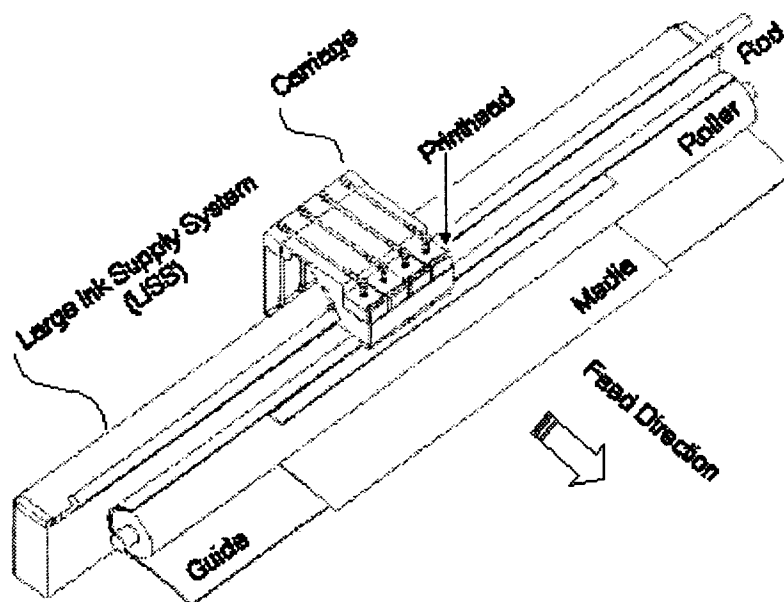


Fig. 2.0: A new design of large ink supply system (LISS) without long tubing



### 3. DISCUSSION FOR LISS

#### 3.1 Back Pressure of System

It's well known that a negative back pressure always needs be kept in the print head to prevent from drilling ink thereof. Note then that the ink reservoir of LISS presented here was open to its environmental air even although it's capped so closed with the belt, sliders, capper and side cover, shown in Figure 3.0. Since the new design used capillary action of ink to automatically and continuously supply the ink usage for the head, it is necessary that the top position of ink reservoir in LISS must be lower than the bottom position of print head, as shown in Figure 4.0. Mathematically, the relation was expressed as equation (3) where  $\delta$  means the distance between the bottom position of print head and the top position of ink reservoir.

$$\delta \begin{cases} > 0, \text{ at least} \\ \cong 2 \sim 4 \text{ cm} \end{cases} \quad (3)$$

Note that the  $\delta$  must be greater than zero in order to keep a negative back pressure of print head. At the same time, values ranging from 2 cm to 4 cm were further preferred. It essentially depended on the character of print head.

#### 3.2 Width of ink reservoir

Since no long tubing system was given in the design of LISS, it required that the ink reservoir was much wider than before. Here, first direction was defined as the feed direction of media so that the direction of width was vertical to it. Moreover, the width must be greater than print width of media in order to allow the print head go covering all margins of media. It might be mathematically described as equation (4) for details. Values of  $W$  were actually depending on the real print width of WFP.

$$W \begin{cases} > \text{Print Width of WFP} \\ \cong 30, 42, 60, \text{ etc... inches} \end{cases} \quad (4)$$

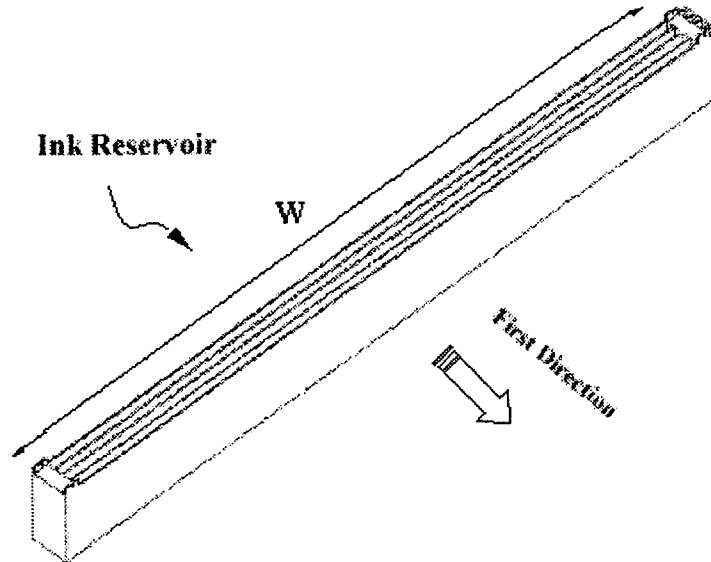


Fig. 5.1: The width of ink reservoir facing the first direction

### 3.3 Depth and height of ink reservoir

Except for the large width of ink reservoir, the depth and height were also characterized so different from before. Figure 5.2 clearly demonstrated the two dimensions. The ink reservoir here was partitioned into four individual chambers where four typical different colors of ink, such as black, cyan, yellow, and magenta, were contained thereof. Each chamber had the depth of  $d$  and the height of  $H$ . Thus the total depth of  $D$  in reservoir was roughly equal to four times of  $d$ . Meanwhile, it's noted that the volume  $V$  of ink reservoir was completely determined by three dimensions, i.e. the width  $W$ , depth  $d$ , and height  $H$ , shown in Figure 5.1 and 5.2. The exact relation could be expressed as  $V=W \times d \times H$ . In addition, the change of height may imply the change of back pressure of print head so that the height should be small enough to avoid large change of back pressure. In general, the amount less than 15 centimeters might be no problem for printing. Therefore, the height could be determined by previous expression if the volume of chamber was given. For example, supposed that one WFP with print width of 54 inches was designed to contain 500-cc ink in each chamber. Next, the width of chamber could be 60 inches and the height of chamber could be 10 centimeters. Thus we could obtain the individual depth  $d$  of 3.3 mm and total depth  $D$  of up to 20 mm as partition thickness counted too. Following the computational rules could easily yield more examples. Generally speaking, it's preferred that the specifications of WFP were holding as expressed in the following relationship (5).

$$\text{General rules, } \begin{cases} H > d \\ H \cong 5 \sim 15 \text{cm} \\ d \geq 0.3 \text{cm} \end{cases}, \quad (5)$$

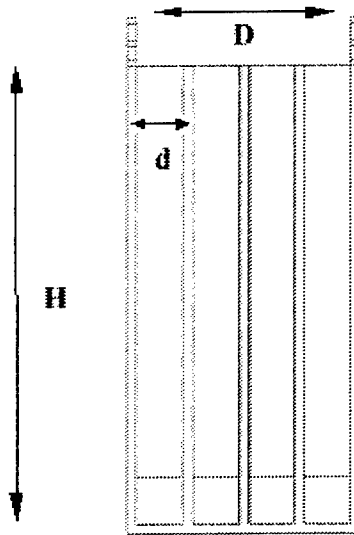


Fig. 5.2: The depth and height of ink reservoir

### 3.4 Clean and fill for ink reservoir

It's also necessarily considered in clean service and filling ink for the LISS. Fill ports were provided in the end of ink reservoir as required to fill ink; of course, each chamber of different ink had such an individual port to fill, respectively. At the same place, the clean ports of chambers were also designed for the service of clean when ink was refilled over time and time. Fill ports and clean ports should certainly at the top and bottom of chamber, respectively. Their corresponding capper and side cover would close them up tightly when the jobs were not requested. One embodiment of LISS was clearly illustrated in Figure 5.3 for details.



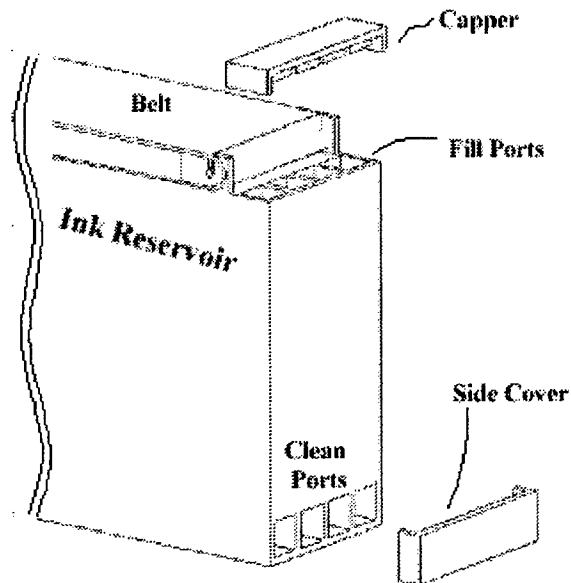


Fig. 5.3: Clean ports and fill ports in the system

### 3.5 Driven belt and slider

It's very necessary to close all chambers of ink reservoir up to prevent the volatility of ink into the air of environment. At the mean time, the ink path via the connector shown in Figure 3.0 and 4.0 should always keep go through from the reservoir to the print head. We noticed here that the print head was not still but movable as any print job been doing. For those purposes, an endless belt was provided in the system. The belt had four holes symbolized as YMCK, for example, to allow the construction of four different ink paths. The distance between any neighboring two was  $d'$  which was approximately equal to the depth  $d$  of chamber. Also it's flexible to form a closed loop that could be synchronically driven with ease by the carriage of print heads.

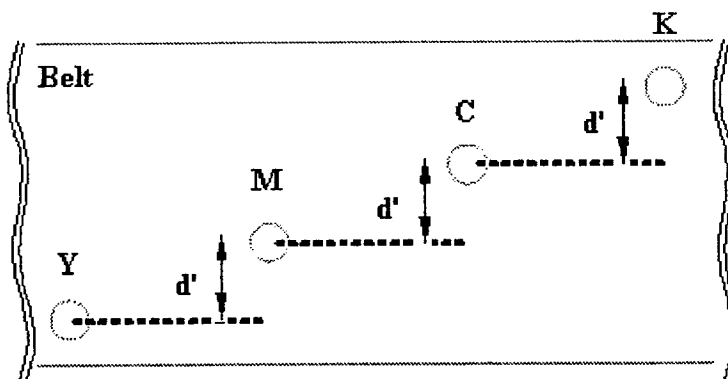


Fig. 5.4: A driven belt with four holes symbolized as Y, M, C, and K

In the two ends of ink reservoir, two sliders were installed as illustrated in Figure 3.0 and 5.3. Their rotation easily allows the belt to move as being driven. Each slider had the width of  $D'$ , shown in Figure 5.5, that approximately equaled the depth  $D$  of reservoir; in addition, the radii of slider could have no limit, but small one in general was preferred.

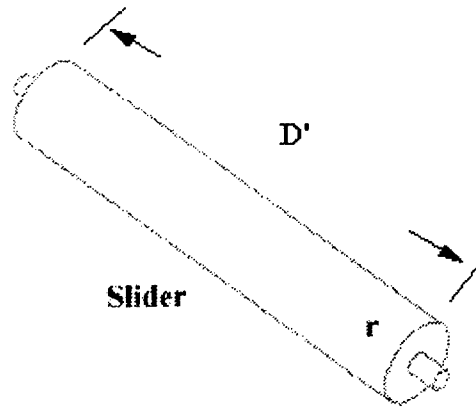


Fig. 5.5: A driven slider with width of  $D'$  and radii of  $r$

## 4. CONCLUSION AND FUTURE WORK

### 4.1 Conclusion

A new method and design of large ink supply system (LISS) without long tubing system for wide format inkjet printer (WFP) was presented in the paper. By replacing long tubing with wider ink reservoir, the traditional pressure loss and pressure vibration could be eliminated a lot. The embodiment of new LISS design was summarized as below-

- **Back pressure of system:** the capillary action was applied to automatically supply the ink for the print head. Therefore, placing the ink reservoir in the position lower than that of the print head could easily set up a negative back pressure. In general, it's preferred to set the distance  $\delta$  as equal to 2 ~ 4 centimeters.
- **Width of ink reservoir:** the width  $W$  of reservoir must be greater than the print width of WFP in order to go through all margins of media. For examples, the values of  $W$  could be 30 inches, 42 inches, and 60 inches for print width of 24 inches, 36 inches, and 52 inches, respectively.
- **Depth and height of ink reservoir:** the depth  $d$  and height  $H$  for each chamber of reservoir could be obtained by the relation  $V=W \times d \times H$  where the symbol  $V$  means the required volume of chamber. Here, general rules were given with  $H > d$ ,  $H \cong 5 \sim 15$  centimeters, and  $d \geq 0.3$  centimeters.
- **Clean and refill:** the capability of cleaning chamber and filling ink over time to time were owned in the system by offering clean ports and fill ports, respectively.
- **Driven belt and slider:** the endless and flexible belt was used to close the ink reservoir up since the volatility of ink in the reservoir must kept as low as possible. In addition, it can synchronically move driven by the print head, via the sliders at the time of movement of the print head.

### 4.2 Future work

Further work may be explored to know how the LISS system performs in the future. What we concern may be several fluid dynamic behaviors of ink in the reservoir, supplying speed of ink, availability of printing speed, Bernoulli effect, and so on. Two approaches of the dynamic simulation/computation and the real experiment of system would be recommended. Their results might further improve the future design presented in the study.

## ACKNOWLEDGEMENTS

This work had been supported by the program MOEA 883NB3110 for wide-format printer project in Optics-Electronics System Labs of Industrial Technology Research Institute in Taiwan<sup>9</sup>. The author really appreciated the support very much.

## REFERENCES

1. C. S. Chan, *Off Board Ink Supply System And Process For Process For Operating An Ink Jet Printer*, US Patent 4831389, Hewlett-Packard Company, 1989.
2. Erickson et al, *Continuous Ink Refill System For Disposable Ink jet Cartridges Having A Predetermined Ink Capacity*, US Patent 5369429, LaserMaster Corporation, 1994.
3. Erickson et al, *Ink Supply Line Support System For A Continuous Ink Refill System For Disposable Ink jet Cartridges*, US Patent 5469201, LaserMaster Corporation, 1995.
4. Murray et al, *Ink jet Printer Incorporating High Volume Ink Reservoirs*, US Patent 5686947, ENCAD Inc., 1997.
5. Sabonis et al, *Ink Source For An Ink Delivery System*, US Patent 5710585, CalComp Inc., 1998.
6. Gragg et al, *Ink Volume Sensing and Replenishing System*, US Patent 5757390, Hewlett-Packard Company, 1998.
7. Robertson et al, *Bulk Ink Delivery System And Method*, US Patent 5751319, Colossal Graphics Inc., 1998.
8. Chin-Tai Chen, "Ink Tank Having Visible Ink Level and End Leg for Ink Supply System of Printer," IS&T NIP15 Conference Proceeding, pp. 59-61, 1999.
9. Chin-Tai Chen, *Design and Method for Large Ink Supply System(Chinese)*, OES-ITRI, Hsinchu, Taiwan R.O.C., 1999.

# The Meniscus Oscillation of Ink Flow Dynamics in Thermal Inkjet Print Head

Ching-Long Chiu, Chien-Wen Wang, Yi-Yung Wu, Yuan-Liang Lan

K200/OES/ITRI  
Bldg. 78, 195-8, Sec. 4, Chung Hsing Road,  
Chutung, Hsinchu 31040, Taiwan, R.O.C.  
Telephone: 886-3-5918438  
Fax: 886-3-5917487  
[H870876@itri.org.tw](mailto:H870876@itri.org.tw)

## Abstract

Meniscus oscillation usually occurs after jetting drops in Drop-on-Demand inkjet head. It is important for ink refill motion. Ink refill motion affects many jetting performance in inkjet technology. Therefore it is very important to understand ink refill motion process for designing inkjet head dimensions and ink properties, such as nozzle diameter, barrier thickness, ink viscosity and surface tension. This meniscus oscillation is a kind of under-damped oscillation. The refill time is defined, as the time required returning to the initialization, the free surface obeys damped oscillation and oscillates between meniscus mounding and recession when the amplitude of the oscillation reduces to dispersion. This study researches several inkjet heads using computational fluid dynamics simulation. CFD solver, such as FLOW\_3D, is used to solve the problem. After calculation, the results are plotted with post-processor such as plot software and output to printers. This paper shows the break-off time, the refill time and operation frequency of the inkjet head.

**Keywords:** Thermal inkjet, Meniscus oscillation, CFD, fluid dynamics.

## 1. Introduction

In recent years, an inkjet printer has become the most dominant output device of PC due to its low cost, high print quality and color printing. However, the print quality is closely related to the ink droplet generation process. The ink droplets are usually jetted by a pressure source generated by thermal bubbles or piezoelectric transducers. When pressure is applied to the bottom of the chamber, the liquid inside the chamber is compressed in order to form droplets jetted from the nozzle which is on the top of the chamber. After the droplet breaks off at the nozzle, the chamber is then refilled with ink again.

The refill motion of ink affects many printing performances. In 1977, Beasley<sup>1</sup> calculated the refill motion for a piezoelectric-driven inkjet head based upon a simple mathematical model, and compared with two experimental results. The analysis reveals that the equivalent length is much shorter than the physical length during the refill, and the calculated refill times agree with the experiments observed. 1981, Kyser<sup>2</sup> introduced a new mathematical model which can be applied to the

known embodiments of jets and shown meniscus oscillation in the refill motion. Saitor<sup>3</sup> used a simplified model under some very simple assumptions to study the refill time and oscillation characteristics. In his study, the neglect of the drop ejection process caused some disagreements with experimental results.

This study uses a Computational Fluid Dynamics (CFD) model to simulate the drop jetting behavior of an inkjet head, and to observe the meniscus oscillation of free surface near the exit of the nozzle plate after drop jetting.

## 2. Numerical description

Simulation of an inkjet head is one of the challenging tasks in CFD. The structure size is quite small, in the order of 10 micrometer, and causes severe truncation errors in the calculated results. Another challenge is keeping track of the free surface between liquid-gas phases. Free surface plays a very important role in our simulations because modeling requires tracking the interface in a more accurate resolution. CFD to be used in this case becomes very complicated, the code has to be capable of capturing the interface accurately. Some packages are capable of dealing the density discontinuity of the free surface by smoothing the interfaces over a few grid cells, but not accounting for the sharp changes. In order to simulate the meniscus oscillation of the inkjet head, the code has to have the capabilities of tracking free surfaces, keeping its sharp changes in one cell distance and evaluating the free surface curvatures.

VOF<sup>4</sup> (Volume Of Fluid) method utilizes a finite difference to represent the free surfaces and interfaces which are arbitrarily oriented with respect to the computational grids. Flow\_3D is a CFD solver using the VOF method in two-phase flow. It can track free surfaces very accurately, and has a certain degree of accuracy in surface sharp. The VOF method is a wide-use two-phase flow model to tackle interface between gas and liquid phases. VOF has been implanted into many packages such as SOLA, SOLA-VOF, NASA-VOF and so on. Today, many inkjet manufacturers and developers use those packages to simulate the jetting process.

VOF has defined a function  $F$  whose value of unity at any point occupied by fluid and zero otherwise. The average value of  $F$  in a cell then represents the fractional volume of the cell which is full of fluid, while a zero value indicates that the cell contains no fluid. There exists a free surface while the value of  $F$  is between zero and one in a cell. Although VOF can locate the free boundary nearly as well as a distribution of marker particle method<sup>4</sup>, and has the advantage of a minimum of stored information, the method is worthless unless an algorithm can be devised for accurately computing the evolution of the  $F$  field. The time dependence of  $F$  is governed by,

$$\frac{\partial F}{\partial t} + u \frac{\partial F}{\partial x} + v \frac{\partial F}{\partial y} = 0 \quad (1)$$

Eqn.1 describes that  $F$  moves with the fluid, and is the partial differential equation of marker particles. Thus, the VOF method provides a simple way to track the free surface.

### 3. Results and Discussions

The simulation of Inkjet head has some difficulties such as the criterion of the bubble formation under the condition of super heating and the free surface capturing. The treatment of the thermal bubble growth of an inkjet head belongs to the super-heating model. So far, there are no commercial packages can deal with super-heating. In fact, it is a big challenge for the inkjet simulation. This study introduces a solid piston to substitute for the thermal bubble, which is a pressure source for inkjet jetting process. Fig.1 is the contour of computational domain for Flow\_3D. The computational domain is an two-dimensional, axisymmetric inkjet head which the radius of the nozzle plate is  $30\ \mu\text{m}$  and  $50\ \mu\text{m}$  in thickness. The height of barrier (or ink supply channel) is  $30\ \mu\text{m}$ . For the convenience of calculation, the solid bottom including a movable piston has been set as  $2\ \mu\text{m}$  thickness. The void region is  $500\ \mu\text{m}$  in length,  $52\ \mu\text{m}$  in radius. The piston has a specified velocity as a function of time shown in Fig.2. The boundary conditions are given as follows: symmetry in central line, wall condition in bottom, continuative condition on the top exit and one atmosphere pressure condition on the right. The viscosity coefficient and surface tension coefficient are  $3.5\ \text{cps}$  and  $50\ \text{dyne/cm}^2$ , respectively.

Fig.3 shows the variation of the interface height in the axial direction for the case in which the nozzle radius is  $30\ \mu\text{m}$  and the height of barrier is  $30\ \mu\text{m}$ . The lowest point of  $27\ \mu\text{sec}$  is the break-off time of the droplet. This figure shows that the oscillation of the free surface. Table 1 shows the computational results for four kinds of nozzle radii and four kinds of barrier heights. From the results, we found that the smaller the nozzle radius, the less the break-off time. Besides, the barrier height does not influence the break-off time if the nozzle radius is kept constant. The results of the refill time show that the larger the nozzle radius, the longer the time.

Some of the results do not obey this tendency, we found that small satellite droplets separated from main droplet or second droplet collide back to the free boundary after jetting. This may cause slightly large oscillation on the free boundary and destroy the natural frequency of damped oscillation. We observe from experiments that there are usually a small amount of ink deposits around the muzzle. Perhaps, the residual ink around the muzzle may have the similar phenomena compared with the simulation results. The reason is that the smaller droplet has opposite momentum while separating from the larger droplet and has larger buoyancy effect.

Table 1 also shows the oscillation frequency of the results. Almost the frequencies are larger than  $2\text{kHz}$  and smaller than  $3\text{kHz}$ . This shows the structure can work at  $3\text{kHz}$ .

From the above results, we use several kinds of nozzle radii and barrier heights to calculate the break-off time of the droplet and the ink refill time. The results show that CFD can use to simulate the behaviors of inkjet heads. Although there are some differences in the break-off time and the refill time, the operation frequencies of the inkjet heads still work about  $2\text{kHz}$ . Here, the design operation frequency is  $3\text{kHz}$ .

### References

1. J. D., Beasley, "Model for Fluid Ejection and Refill in an Impulse Drive Jet," *Photographic Science and Engineering* **21**, pp. 78-82, 1977.
2. E. L., Kyser, L. F. Collins, and N. Herbert, "Design of an Impulse Ink Jet," *Journal of Applied Photographic Engineering* **7**, pp.73-79, 1981.
3. K. Saitoh, "A Simple Model of Ink Refill Motion in Inkjet Printing," 7<sup>th</sup> IS&T's International Congress on Advances in Non-Impact Printing Technologies, Portland, Oregon, Volume II, pp.120-123, 1991.
4. C. W. Hirt, and B. D. Nichols, "Volume of Fluid (VOF) Method for the Dynamics of Free Boundaries," *Journal of Computational Physics* **39**, pp.201-225, 1981.

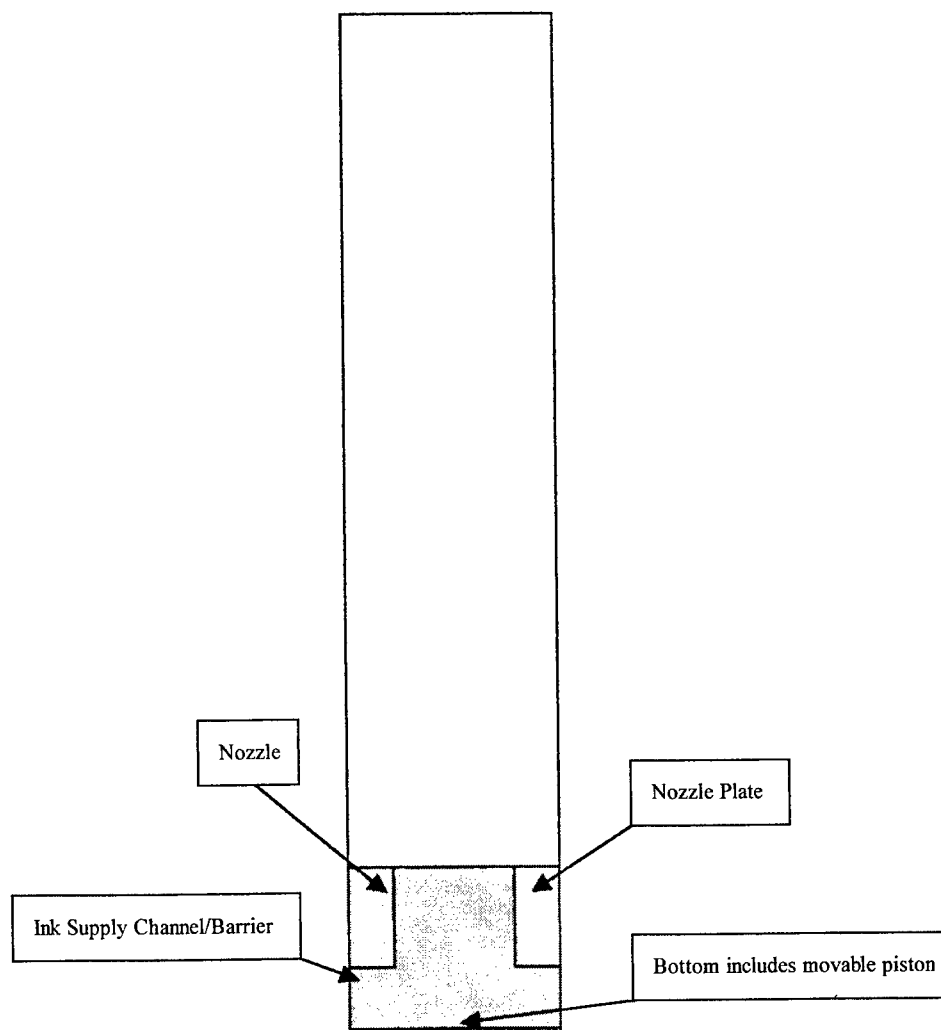


Fig.1 2D computational domain

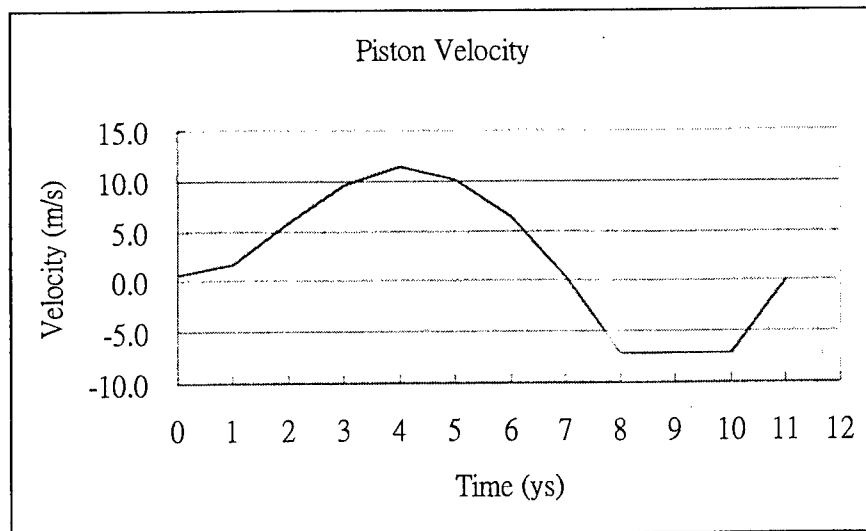


Fig. 2 The piston velocity profile

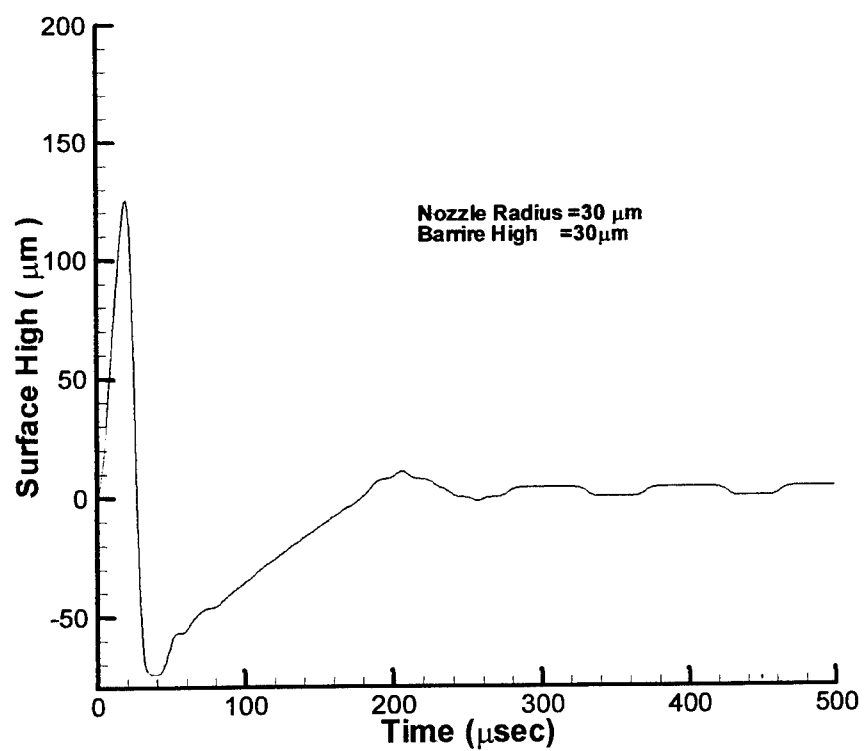


Fig.3 The surface height variation



Break-off Time, $\mu$ sec				
Nozzle Radius, $\mu$ m	26	28	30	32
Barrier Height, $\mu$ m				
25	23	24	29	31
28	22	24	29	28
30	23	24	27	28
34	24	25	27	28
Refill Time, $\mu$ sec				
Nozzle Radius, $\mu$ m	26	28	30	32
Barrier Height, $\mu$ m				
25	497	331	335	452
28	320	420	366	497
30	327	355	385	361
34	344	355	385	447
Frequency, kHz				
Nozzle Radius, $\mu$ m	26	28	30	32
Barrier Height, $\mu$ m				
25	2.01	3.02	2.99	2.21
28	3.13	2.38	2.73	2.01
30	3.06	2.82	2.60	2.77
34	2.91	2.82	2.60	2.24

Table 1 The break-off time, refill time, and operation frequency of several nozzle radii and barrier heights

# Measurement of Contrast Ratios for 3D Display

Kuo-Chung Huang<sup>1</sup>, Chao-Hsu Tsai<sup>1&2</sup>, Kuen Lee<sup>1</sup>, Wen-Jean Hsueh<sup>1</sup>

<sup>1</sup>ITRI/OES, 195-8, Sec. 4, Chung Hsing Rd. Chutung, Hsinchu 310 Taiwan, R.O.C

<sup>2</sup>National Taiwan University, Institute of Applied Mechanics

**Keyword:** contrast, crosstalk, stereoscopic, autostereoscopic, 3D, display

## ABSTRACT

3D image display devices have wide applications in medical and entertainment areas. Binocular (stereoscopic) imaging without glasses, especially spatial-multiplexed displays such as lenticular display, barrier strip display, and single-lens stereoscopic display, is one of the most powerful and popular ways for life-like presentation of our three-dimensional environments.

The definition and relationship of the image contrast and viewer crosstalk are reviewed and clarified. They are measured and compared on three different types of 3D display systems, including shutter-glasses stereoscopic display, image splitter autostereoscopic display and dual-panel autostereoscopic display.

From the contrast point of view, high-quality three-dimensional perception results from a combination of high image contrast and low crosstalk. Same as a conventional two-dimensional display, high image contrast is also required for a 3D display to present a satisfactory image to either eye of the viewer. Yet, there is an extra requirement for a 3D display. The viewer crosstalk must be low enough for the viewer's one eye to neglect the ghost image from the neighboring viewing zone of the other eye.

The interesting fact is that there are conflicts between these two factors to generate satisfactory 3D effects. As a characteristic of the display system, the system crosstalk will confine a content provider within a certain range of image contrast to present satisfactory 3D pictures or videos to the viewer.

## 1. INTRODUCTION

The elementary cues of 3D vision from a stereoscopic display system are from a series of images with lateral disparity. When two of the images with proper disparity are fed into a viewer's two eyes, the viewer obtains his first cue of 3D perception - binocular vision. The viewer will obtain his second cue from the display while he moves his head laterally and sees different images from the corresponding aspect. This is called the motion parallax. Therefore, an ideal stereoscopic display has to make efforts to present corresponding images at different viewing angles to let the viewer obtain binocular parallax and motion parallax.

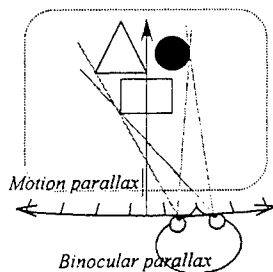


Figure 1. Stereopsis – a combination of the binocular parallax and motion parallax.

Some of the earlier researches in the physiological optic area pointed out that the quality of the 3D perception depends on the contrast of the composing 2D images<sup>1,2,3</sup>. The higher the image contrast is, the better the 3D perception will be. But probably they did not have a more practical display system as today, usually they obtained their conclusion about how monocular contrast affects stereo acuity by using “mirror stereoscopic” experimental setups.

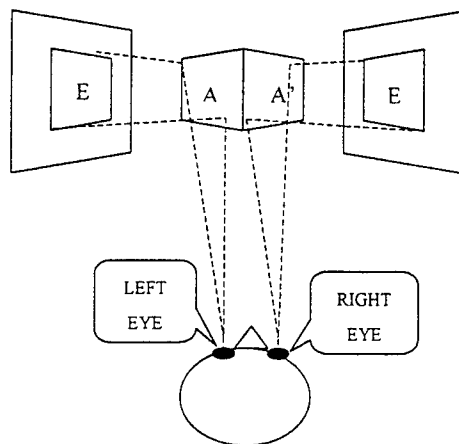


Figure 2. Mirror stereoscopic display. From the reflection of the two mirrors (A and A'), the two pictures are presented into the two eyes of the viewer, without any “crosstalk” problem.

In the recent years, a new important factor, which is called “ghost image”, arises during the development of the stereoscopic display technologies. The “ghost image” results from the crosstalk between the viewing zones for the left and right eyes of the viewer. The amount of the crosstalk reduces the ability of the viewer to fuse the two eye views into a single, 3 dimensional one.

Therefore, we would like to clarify the difference between the image contrast (this affects the monocular image quality) and viewer crosstalk in this paper. Furthermore, the relationship between the image contrast, system crosstalk and viewer crosstalk will be discussed.

## 2. BACKGROUND OF 3D DISPLAY

Let's take an overview of 3D display technology<sup>8,9</sup>. There are various ways of creating a three dimensional display. The practical methods that meet the necessary attributes fall into three categories: Hologram displays, spatially multiplexed displays and time multiplexed displays.

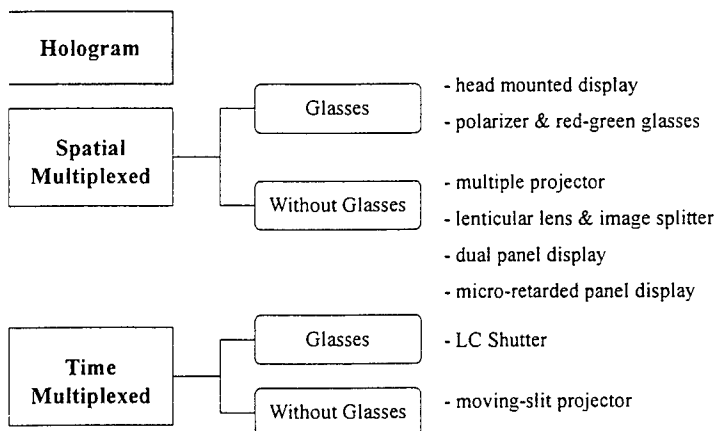


Figure 3. Three categories of 3D display.

In this paper, we will evaluate the interaction between the image contrast and viewer crosstalk by experiments on four different stereoscopic display systems, representing different categories of stereoscopic display respectively. The four display systems are described as following.

## 2.1 SHUTTER-GLASSES STEREOSCOPIC DISPLAY

A widely applied display method using time multiplexed method to present binocular visions<sup>4</sup>.

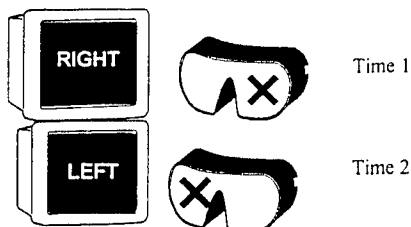


Figure 4. Shutter Glasses: Time 1 to show one eye's image, and Time 2 to show the other.

## 2.2 IMAGE SPLITTER AUTOSTEREOSCOPIC DISPLAY

An image splitter autostereoscopic display, or parallax barrier autostereoscopic display, is made by placing a barrier-strip plate in front of an image display panel at a pre-designed distance. Each image to be presented on such a display is vertically interlaced from two images of different parallax, one for the right eye and the other for the left eye.

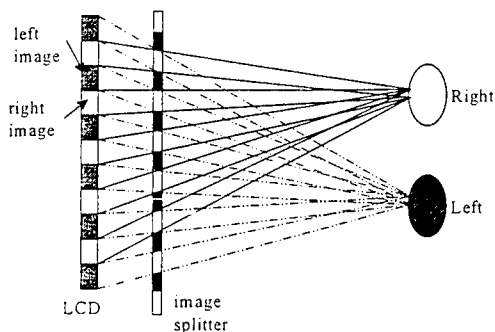


Figure 5. Image splitter autostereoscopic display.

## 2.3 DUAL-PANEL AUTOSTEREOSCOPIC DISPLAY

Mahler<sup>6</sup> used two pieces of transparencies in front of large lenses and a beam splitter to combine the two images, and now we use LCD panels to replace the light sources and transparencies.

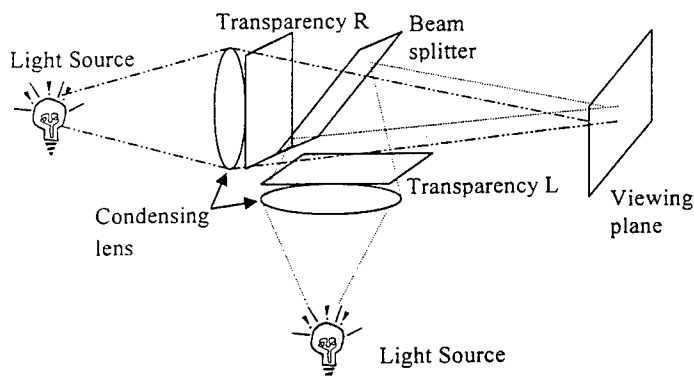


Figure 6. A dual-panel autostereoscopic display adapted from Mahler's "Photoplastikon".

### 3. THE DEFINITIONS

Before we give a more precise definition to the various contrasts and crosstalks<sup>5,6</sup>, we will introduce “co-location point” first.

While watching a 3D object in the real world, there will be two different 2D images (with disparity) formed on the retinas of the two eyes of the viewer. There will be various discrepancies between these two images because they have disparity. Like in figure 7, at a specific location on the display screen one will find a part of one image (e.g., the right edge of a rod in the right-eye image), and an adjacent part of the other image (e.g., the background in the left-eye image). The representing position  $(x_o, y_o)$  of these two parts on the two images is defined as “co-location point” in this paper. Please be noted that, the definition of the “co-location point” is different from that of the “corresponding point”, which is point A and A' in the right-eye image and left-eye image respectively.

It is not expected to see ghost images on an ideal stereoscopic display. However, most of people will notice and be bothered by ghost images on stereoscopic displays, especially on an autostereoscopic display. The ghost image is usually due to the incapability of the display system to totally clear up the light leakage from the co-location point of the other eye's image.

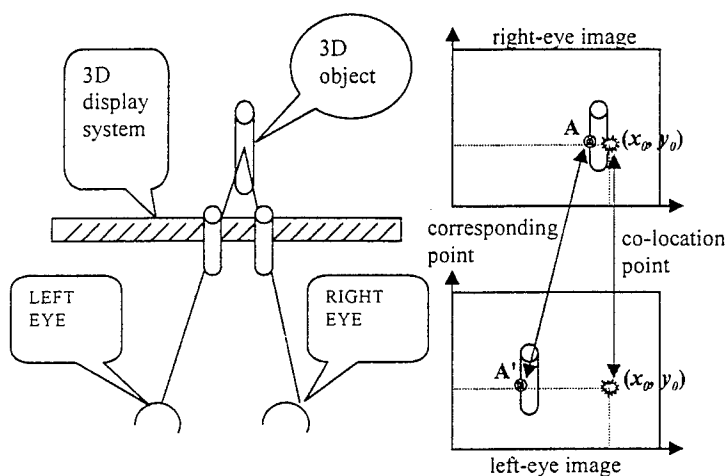
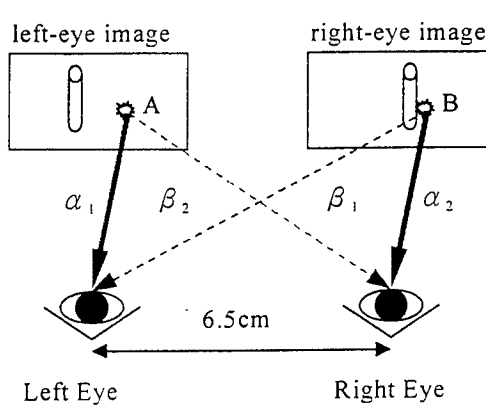
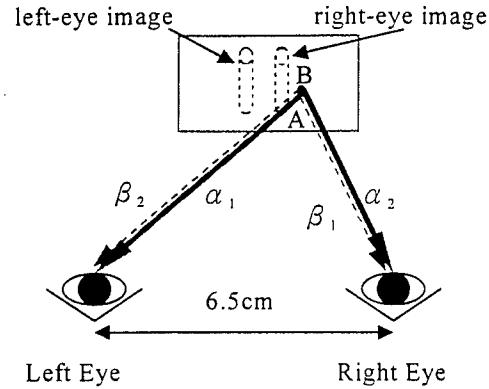


Figure 7. A diagram for explaining a stereoscopic display system to show a 3D object.

Now please refer to Figure 8. Numbers A and B are the luminance values at the specified points on the screen, and  $\alpha_1, \alpha_2, \beta_1, \beta_2$  are stereoscopic display system parameters. In Fig. 8a, A and B represent the luminance value of the co-location points on left-eye and right-eye images, respectively, in a dual panel autostereoscopic display. In Fig. 8b, A and B represent the luminance value of the co-location points on interlaced left-eye and right-eye images, respectively, in the other three stereoscopic displays.  $\alpha_1, \alpha_2$  describe the percentage part of the left-eye image should be observed at the left-eye position and the percentage part of the right-eye image should be observed at the right-eye position, respectively.  $\beta_1$  and  $\beta_2$  describe the percentage part of the left-eye image leaked to the left-eye position and the percentage part of the right-eye image leaked to the right-eye position, respectively.



(Fig.8a)



(Fig.8b)

Figure 8. Diagram for the parameters for the definitions. Fig.8a is for dual panel system examples, Fig.8b is for single panel system examples.

### 3.1 CO-LOCATION IMAGE CONTRAST

When a 3D content is displayed on a stereoscopic display, the co-location image contrast of a specific point on the screen is “defined as” the luminance ratio of the co-location points of the two images (left-eye and right-eye) at that specific position. Carefully review the above definition, the co-location image contrast is actually affected by two different factors, the content itself and the image source property (e.g., the panel contrast of LCD or CRT).

For a left-eye case, the co-location image contrast can be defined as the luminance ratio  $B/A$ .

Note that the co-location image contrast is different from the well-known image contrast which is defined as the luminance ratio of the point with maximum luminance to the point with minimum luminance for the entire 2D image.

It is easy to prove that if the image contrast is adjusted to a new value, the relationship between the new co-location image contrast and the image contrast will follow the equation

$$v'/u' = (v + c)/(u + c) \text{ ----- (1)}$$

where  $v'/u'$  is the co-location image contrast before adjustment,  $v/u$  is the co-location image contrast after adjustment;  $c$  is a constant equal to  $L_{\min} \times (C_r - C_r')/(C_r' - 1)$ , where  $C_r$  and  $C_r'$  are the image contrast before and after adjustment as defined above,  $L_{\min}$  is the minimum luminance in the image before adjustment. Taking a detail look at the relationship, we can find out that the co-location image contrast for each specific co-location pair will increase when the 2D image contrast of each single-view image gets higher.

### 3.2 SYSTEM CROSSTALK

This value is used to evaluate the optical performance of the stereoscopic display system, and is independent of the content. From Figure 8, for a left-eye case, the system crosstalk can be defined as  $\beta_2/\alpha_1$ , describing the degree of the unexpected leaking image from the other eye.

### 3.3 VIEWER CROSSTALK

In the former two paragraphs, the two important factors affecting the ghost image a stereoscopic viewer will sense are defined. Here the authors will further quantize the “ghost image”, or the viewer crosstalk, which is measured at the viewer's side. The viewer crosstalk is defined as the ratio of the luminance of unwanted “ghost” image to the luminance of the correct information received by the viewer's eyes<sup>6</sup>. Referring to Figure 8, the viewer crosstalk for the viewer's left eye can be defined as  $B\beta_2/\alpha_1$ .

#### 4. MEASUREMENT OF SYSTEM CROSSTALK

From the above definitions, a simple relationship can be found. The viewer crosstalk  $B\beta_2/A\alpha_1$  can be written as the product of  $(B/A)$  and  $(\beta_2/\alpha_1)$ , or

$$\text{viewer crosstalk} = \text{co-location image contrast} * \text{system crosstalk} \text{ ----- (2)}$$

For the simplicity of measurement, the viewer crosstalk can be expressed as  $(B\alpha_2/A\alpha_1) * (\beta_2/\alpha_2)$ , too. We measured the different stereoscopic systems in our laboratory and verified the above relationships.

For the purpose of quantification, different levels of gray-scaled patterns are applied for measurement. For every display system, the quantities  $B\beta_2/A\alpha_1$  and  $B\alpha_2/A\alpha_1$  are measured with different gray scales of patterns. All the measurements are done by Minolta CS-100 luminance meter.

##### 4.1 SHUTTER-GLASSES STEREOSCOPIC DISPLAY

This experimental system is specified as the following. The display device is a CRT display with P 22 phosphors, SVGA resolution, and refresh rate 120 Hz. The shutter-glasses is purchased from APEC Inc., Taiwan, model number VR97. The shutter-glasses is fixed in front of the CRT (with a distance of 60 cm and the luminance meter is positioned behind the glasses. The result is,

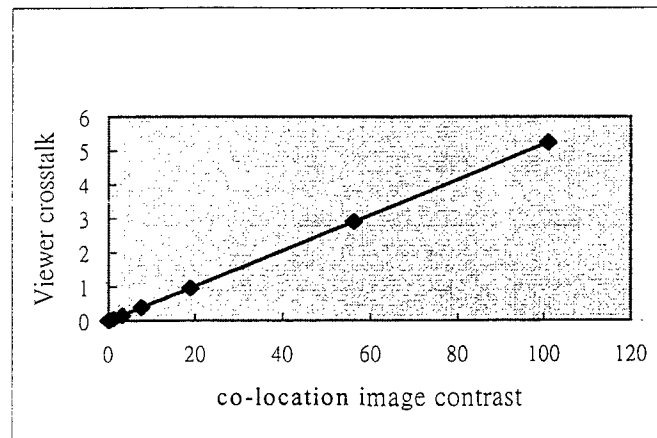


Figure 9. Viewer crosstalk of shutter glasses stereoscopic display.

##### 4.2 IMAGE SPLITTER AUTOSTEREOSCOPIC DISPLAY

The image splitter autostereoscopic display specifications are, the display device is Sanyo's 15" LCD with double image splitter, with luminance 400 cd/m<sup>2</sup>, proper viewing distance 23", and XGA resolution. The luminance meter is set at the best observation position. The result is,

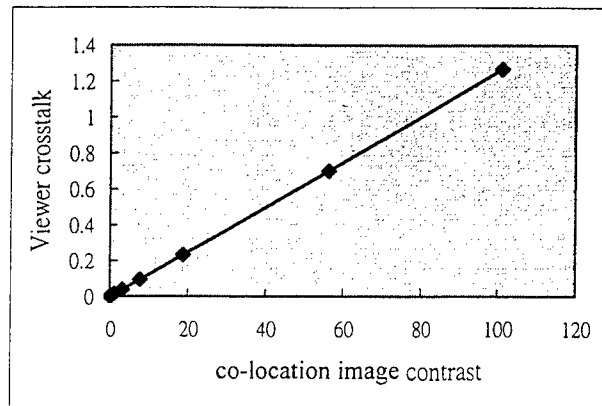


Figure 10. Viewer crosstalk of image splitter autostereoscopic display.

#### 4.3 DUAL-PANEL AUTOSTEREOSCOPIC DISPLAY

This experimental system is a little bit more complicated, please refer to figure 6. The LCD panels are Philips' 15.1" XGA panels, screen luminance is 7 cd/m<sup>2</sup>, and the most proper viewing distance is 60 cm. The backlight brightness and panel contrast are adjusted to the same level for the two LCDs. The result is,

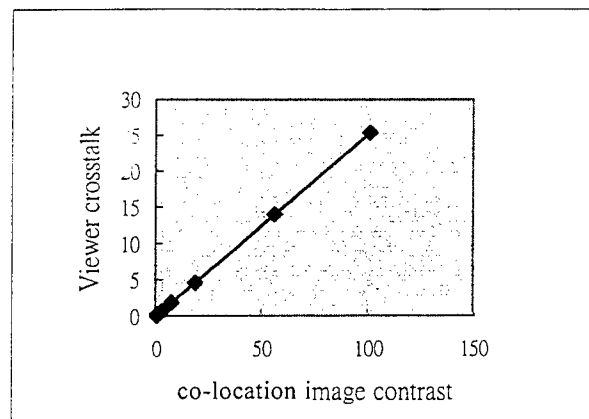


Figure 11. Viewer crosstalk of dual-panel autostereoscopic display.

#### 5. DISCUSSION

Observing the above measurement results, it is obvious that the slopes  $\beta_2/\alpha_2$  in the plots are all basically constants for every case. Refer to equation 2, the parameter  $\beta_2/\alpha_2$  differs from the system crosstalk of the stereoscopic systems ( $\beta_2/\alpha_1$ ) by a constant  $\alpha_2/\alpha_1$ . For a practical system, it is reasonable to believe that  $\alpha_1$  and  $\alpha_2$  are close to each other. In such a case, the slopes will be close to the system crosstalk of the stereoscopic system and can also be used to characterize the stereoscopic system. The slopes  $\beta_2/\alpha_2$  for the different 3D systems can be summarized as following.

system	shutter-glasses	dual-panel	image splitter
--------	-----------------	------------	----------------



Slope	0.058	0.28	0.014
-------	-------	------	-------

The system crosstalk (or the slope  $\beta_2/\alpha_2$ ) is an index of optical performance for a stereoscopic system. The smaller the system crosstalk is, the less possible the viewer is tending to see ghost image. It is independent of the content, such as the co-location image contrast, single-view image contrast or the image brightness. In this paper, the system crosstalk is measured from the centers of the viewing zones (of our autostereoscopic displays to acquire a best result. Yet practically the viewer's eyes are not necessarily fell in the best position while watching autostereoscopic images. In fact, the system crosstalk will increase as the position of measurement leaves the center of the viewing zone. Therefore, the variation of the value of the system crosstalk ratio may indicate the range in which one can obtain 3D vision with good quality on a stereoscopic display.

However, the parameter that finally determines the level of ghost image the viewer will see is the viewer's crosstalk instead of the system crosstalk. It is clear that the co-location image contrast is defined for interpretation for the relationship of the system crosstalk and the viewer crosstalk. Therefore, it is a localized property of the image, i.e., the value of the co-location image contrast is decided from a co-location point of the left-eye and right-eye images.

For a 2D display, the higher the image contrast is, the better the visual quality will be. But this is not necessarily correct for a stereoscopic display. According to equation 2, the viewer's crosstalk is proportional to the co-location contrast. That means as the co-location contrast gets higher and higher, the ghost image phenomena will become worse and worse. It has been mentioned in the previous paragraph that the co-location contrast and the single-view image contrast has a positive relationship, i.e. increase in one value will also induce increase in the other, and vice versa. Under this fact, although the image contrast is indeed proportional to the visual quality for monocular vision of the viewer, a higher image contrast will on the contrary cause more crosstalk for a stereoscopic system. It is because the better image contrast will be easier to induce "ghost image" due to the non-infinitesimal value of the system crosstalk. In another word, the image contrast could be a conflict factor with the system crosstalk in a stereoscopic display system. The "viewer crosstalk" is then an overall evaluation for the ghost image, and is easy to be interpreted due to the principle of binocular 3D display<sup>4</sup>.

## 6. CONCLUSION

In this paper, a simple method to measure the system crosstalk is pointed out and a relationship between the image contrast and viewer crosstalk is established<sup>10</sup>. Our result can get the following conclusions:

1. When the system crosstalk is a constant, the image contrast is not the higher the better. This result does not depend on any specific system. At least, from the four different stereoscopic displays the same conclusion is derived.
2. System crosstalk is a evaluation of the performance of a stereoscopic display. Stereo contrast is equal to the product of the image contrast and system crosstalk.

In many 3D display systems viewer' crosstalk is an important issue for good performance, especial in autostereoscopic display systems. For an autostereoscopic display system, system crosstalk is not always the same when observer is at different position in front of the screen, even there is a tracking system. Therefore, the viewing angle of the system can be decided by the system crosstalk measurement at different positions. On the other hand, the viewer crosstalk describing the seriousness of ghost image is an overall result of the image contrast and the system crosstalk. Image processing method can be applied to decrease image contrast to decrease viewer crosstalk.

There are still issues should be studied further. For example, the maximum viewer's crosstalk allowed for a viewer to obtain good 3D perception. This paper probes criteria for a good 3D display only from the luminance point of view, other factors like spatial frequency and cross- or uncross- disparity are not considered yet<sup>7</sup>.

## ACKNOWLEDGEMENTS

Research on the autostereoscopic display technology has been supported by the Ministry of Economic Affairs of Taiwan, R.O.C. and Opto-Electronics & Systems Lab of ITRI. The authors gratefully acknowledge the support of engineers in the Stereo Lab, including Dr. Fang-Chuan Ho and Dr. Ruan-Ywan, Tsai.

## REFERENCES

1. D Lynn Halpern ,Randolph R Blake, "How contrast affects stereoacuity ",*Percetion*,1988,volume 17,number 4,page 483-495
- 2.Blake R, Cormack R H,1979 "Does contrast disparityalone generate stereopsis?" *Vision Research* **19** 913-915
- 3.Friby J P, Maybew J E W, 1978 "Contrast sensitivity function for stereopsis " *Perception* **7** 423-429
- 4 Y. Yeh and L.D. Silverstein, "Limits of fusion and depth judgement in stereoscopic color displays", *Human Factors*, v.32, pp. 45-60, 1990.
5. Peter G.J.Barten ".Contrast Sensitivity of the HUMAN EYE and Its Effect on Image " *SPIE OPTICAL ENGINEERING PRESS*.
6. S.A. Benton, T.E. Slowe, A.B. Kropp and S.L. Smith, "Micropolarizer-based multiple-viewer autostereoscopic display", *Proc. SPIE* 3639, *SPIE Symposium on Stereoscopic Displays and Applications X*, pp.76-83, Feb. 1999.
- 7 Y. Yeh and L.D. Silverstein, "Human Factor for stereoscopic color displays", *SID 91 Digest*, pp. 826-829.
8. J.R. Moore, N.A. Dodgson, A.R.L. Travis and S.R. Lang, "Time-multiplexed color autostereoscopic display", *Proc. SPIE* 2653, *SPIE Symposium on Stereoscopic Displays and Applications VII*, pp. 10-19, Feb. 1996.
9. S.A. Benton, "The second generation of the MIT holographic video system," in: J. Tsujiuchi, J. Hamasaki, and M. Wada, eds., *Proc. of the TAO (Telecommunications Advancement Organization of Japan) First International Symposium on Three Dimensional Image Communication Technologies*, pp. S-3-1-1 to -6, Dec. 1993..
- 10.H. Morishama, H. Nose, N. Taniguchi ,K. Inoguchi, S. Matsumura "An Eyeglass-Free Rear-Cross-Lenticular 3-D Display" , *SID DIGEST* 98

# Spatial long-range modulation of contrast discrimination

Chien-Chung Chen<sup>a\*</sup> & Christopher W. Tyler<sup>b</sup>

<sup>a</sup> Ophthalmology Department, University of British Columbia, Vancouver, British Columbia, Canada

<sup>b</sup> Smith-Kettlewell Eye Research Institute, San Francisco, California, USA

## ABSTRACT

Contrast discrimination is an important type of information for establishing image quality metrics based on human vision. We used a dual-masking paradigm to study how contrast discrimination can be influenced by the presence of adjacent stimuli. In a dual masking paradigm, the observer's task is to detect a target superimposed on a pedestal in the presence of flankers. The flankers (1) reduce the target threshold at zero pedestal contrast; (2) reduce the size of pedestal facilitation at low pedestal contrasts; and (3) shift the TvC (Target threshold vs. pedestal contrast) function horizontally to the left on a log-log plot at high pedestal contrasts. The horizontal shift at high pedestal contrasts suggests that the flanker effect is a multiplicative factor that cannot be explained by previous models of contrast discrimination. We extended a divisive inhibition model of contrast discrimination by implementing the flanker effect as a multiplicative sensitivity modulation factor that account for the data well.

**Keywords:** image quality assessment, human vision, dual masking, divisive inhibition

## 1. Introduction

### 1.1 Contrast discrimination and the divisive inhibition models

A good image quality metrics should assess the quality of an image in a way that is consistent with human visual experience. Thus, much effort has been expended in incorporating human visual psychophysics data or models based on the human visual system into image quality metrics<sup>5,19,22,22,24</sup>. The most relevant human psychophysical data for image quality assessment are contrast discrimination thresholds. In a typical contrast discrimination experiment, the task of an observer is to detect a periodic pattern (target) superimposed on another periodic pattern (pedestal) with the same spatiotemporal properties except contrast. The contrast discrimination threshold, or target threshold, is defined as the target contrast that allows the observer to tell the difference between pedestal-alone and pedestal-plus-target with certain percentage of correctness.

A typical result of contrast discrimination experiments is the "dipper" shaped *target contrast vs. pedestal contrast* (TvC) function<sup>6,7,10,16,25</sup>. That is, as pedestal contrast increases, the target contrast threshold first decreases (facilitation) and then increases (masking) as shown in figure 1a. The TvC function reflects the contrast response characteristics in the visual system. As shown in Figure 1, in order to be detected, the target has to have enough contrast to increase the response to the pedestal alone by a certain amount, defined as one unit. Suppose that the response function is accelerating near a pedestal contrast (e.g.,  $C_1$  in Figure 1), it would require less target contrast ( $\Delta C_1$ ) than the unmasked threshold ( $C_{t0}$ , the threshold measured without the pedestal) to increase the response by one unit. On the other hand, when the response function is decelerating near a pedestal contrast (e.g.,  $C_2$  in Figure 1), it would take greater target contrast ( $\Delta C_1$ ) to increase the response by the same amount. Thus, the target threshold at a pedestal contrast is inversely proportional to the slope of the response function at that pedestal contrast.

Currently, the most popular model of contrast response functions is the divisive inhibition model, also called the contrast normalization model<sup>1,3,4,6,8,16,17,19,23</sup>. Different variations of this model have been used in several image quality metrics. All the variations of the divisive inhibition models share the following common features:

---

\* Correspondence: E-mail [chen@ski.org](mailto:chen@ski.org); Telephone: 1(604)6097225

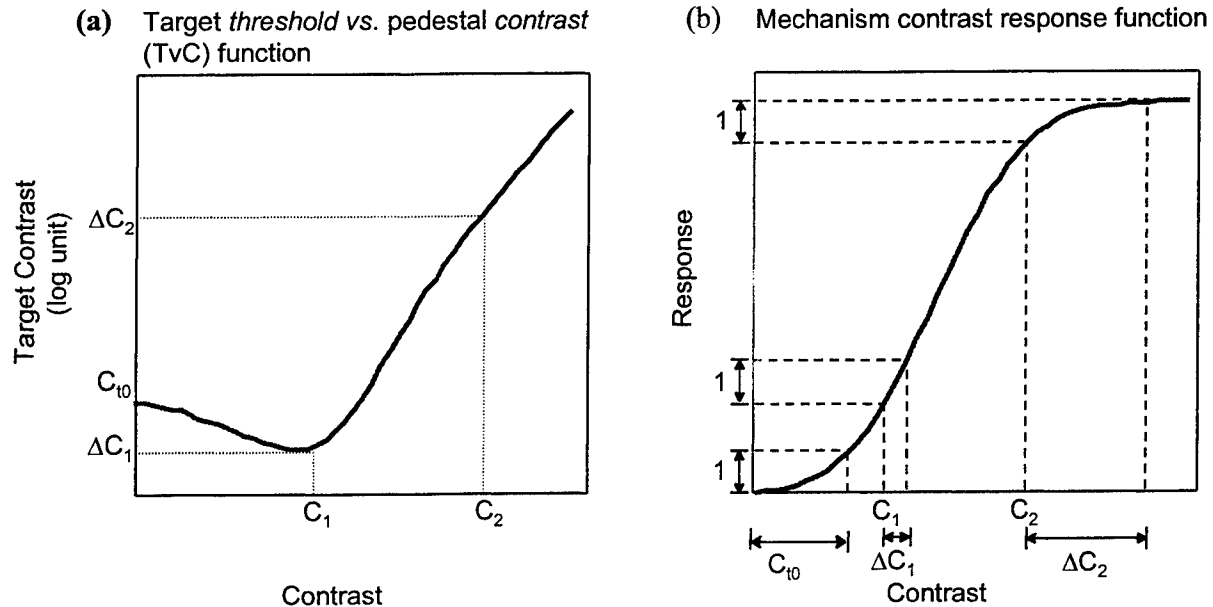


Figure 1. The relationship between the target threshold vs. pedestal contrast (TvC) function (left panel) and the contrast response function (right panel).  $C_{t0}$  is the target absolute (unmasked) threshold.  $\Delta C_1$  and  $\Delta C_2$  are target thresholds measured with the presence of pedestal contrast  $C_1$  and  $C_2$  respectively. At threshold, the target increases the response to pedestal alone by one unit. As a result, the target threshold is inversely proportion to the slope of the response function.

- (1) Linear filters. The input images are processed by a band of linear filters defined by wavelet functions such as Gabor or difference-of-Gaussian. Those linear filters have a limited spatial extent and are localized in both space and the Fourier domain. The output of each linear filter can be simplified as the contrast of the Fourier component to which it is tuned weighted by a constant<sup>4</sup>.
- (2) Divisive inhibition. Each linear filter is followed by a nonlinear response operator. Each nonlinear operator takes two inputs: i) the excitatory input fed directly from the corresponding linear filters and ii) the divisive inhibitory input which is pooled from the outputs of all relevant linear filters (normalization pools). The response of each nonlinear operator is the excitatory input raised by a power and then divided by the divisive inhibitory input plus a constant.
- (3) Decision. In a two-alternative forced-choice contrast discrimination experiment, the observer compares the nonlinear operator responses to the pedestal-alone and to the pedestal-plus-target. The decision is based on the operator that shows the greatest difference between the responses. The target is at threshold when the difference reaches a criterion.

## 1.2 Flanker effect

Notice that the divisive inhibition models are based on localized mechanisms. These models were designed to account for the contrast discrimination experiments where only the targets and the pedestals were presented and they did not take the spatial context of the stimuli into account. However, recent studies have shown that spatial context does affect human visual behavior. For instance, Polat & Sagi<sup>14,15</sup> measured detection thresholds for a target Gabor pattern at the fovea flanked by two other high contrast Gabor patterns (flankers). The target threshold decreased up to about 50% of the absolute threshold (facilitation) when a pair of collinear flankers (where the flanker had the same orientation as the target) presented at about three units of target wavelength away. Conversely, flankers with an orientation that was orthogonal to the target had no effect on target detection. This control result establishes that the effects of the flankers are not generic attention or uncertainty effects but are local or long-range interactions specific to the receptive field and orientation selectivity.

Current divisive inhibition models can deal with the flanker effect in two ways: (1) Increase the size of the linear filters such that the target and the flankers can be covered by the same filter. Thus, the presence of flankers can produce a response in the target mechanism and in turn affect the target threshold<sup>18</sup>. (2) Increase the extent of the contrast normalization pools<sup>17</sup>. Thus, the presence of flankers just adds an extra term in the contrast normalization signal. In either way, the effect of the flanker is an additive term to the effect of the target as proposed both by Snowden & Hammet<sup>17</sup> and Solomon et al.<sup>18</sup>. However, as we will show in this study, the flanker effect is not additive but multiplicative. Our experiment shows that the current divisive inhibition models cannot account for contrast discrimination thresholds in the presence of flankers.

### 1.3 Dual Masking Paradigm

Consider a dual masking experiment in which the task of the observer is to detect a target pattern superimposed on a pedestal (first masker) and in the presence of flankers (second masker). When the pedestal and the target have the same spatial temporal properties except contrast, this dual making experiment is equivalent to a contrast discrimination experiment conducted in the presence of flankers. Thus, if we systematically measure the target contrast on various pedestal contrasts, we can obtain a contrast discrimination TvC function under the influence of flankers.

Figure 2 shows the prediction of current divisive inhibition model to the flanker effect. The solid curve is the contrast discrimination TvC function measured without the presence of flankers. When the flankers are presented, according to the divisive inhibition models, their effects are added to that of the pedestal. Suppose that the flanker contrast is a constant through the experiment. The flanker effect should also be a constant. On the other hand, the effect of pedestal increases with its contrast. Thus, comparing to the pedestal effect, the flanker effect should be less and less salient as the pedestal contrast increases. Therefore, the divisive inhibition model should predict that the TvC functions with the presence of the flankers should converge to the TvC functions without the flankers as pedestal contrast increases. A detailed discussion of the divisive inhibition model predictions can found in Snowden & Hammet<sup>17</sup> in a slightly different context. The dashed curve in Figure 2 illustrates such prediction. The experiment report below will show this prediction fails to capture the characteristics of the data.

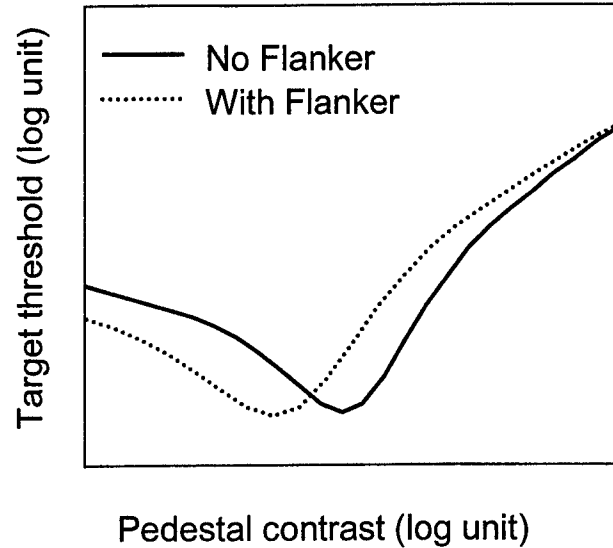


Figure 2. The divisive inhibition model prediction on flanker effects. The solid curve is the TvC function measured without the presence of flankers and dotted curve is that measured with the presence of flankers. The divisive inhibition model predicts that the two TvC functions are getting closer to each other as pedestal contrast increases.

The dashed curve in Figure 2 illustrates such prediction. The experiment report below will show this prediction fails to capture the characteristics of the data.

## 2. Methods

### 2.1 Apparatus

The stimuli were presented on two SONY CPD-1425 monitors each driven by a Radius PrecisionColor graphic board. A Macintosh Quadra Pro computer controlled the graphic boards. The resolution of the monitor was 640 horizontal by 480 vertical pixels. At the viewing distance we used (128 cm), there were 60 pixels per degree. The viewing field was then 10.7° (H) by 8° (V). The refresh rate of the monitor was 60 Hz. We used the LightMouse photometer (Tyler, 1997) to measure the input-output intensity function of the monitor. This information allowed us to compute linear lookup table settings. The mean luminance of the monitor was set at 35 cd/m<sup>2</sup>.

### 2.2 Stimuli

The target, the pedestal and the flankers were all vertical Gabor patches defined by the equation

$$G(x,y) = BG + BG * C * \cos(2\pi kx) * \exp(-x^2/2\sigma^2) * \exp(-(y-u_y)^2/2\sigma^2)$$

where BG was the mean luminance, C, ranged from 0 to 1, was the contrast of the pattern, k was the spatial frequency,  $\sigma$  was the scale parameter (standard deviation) of the Gaussian envelope and  $u_y$  was the displacement of the pattern. All patterns had a spatial frequency (k) of 4 cycles per degree and a scale parameter ( $\sigma$ )  $0.1768^\circ$ . The target and the pedestal were centered at the fixation point, therefore the displacement  $u_y$  was zero. The two flankers were placed at the top and below the target. The displacement was  $0.75^\circ$ . The stimuli were presented concurrently. The temporal waveform of the stimulus was a pulse with duration of 100 msec. The contrast of the patterns is presented in decibels (dB), which is 20 times log base 10 of the linear contrast.

### 2.3 Procedures

We used a temporal two-alternative forced-choice paradigm to measure the target threshold. In each trial, the pedestal and the flankers were presented at both intervals. The target was randomly presented in either one of the intervals. The task of the observer was to determine which interval contains the target. We used the Psi adaptive threshold seeking algorithm<sup>9</sup> to measure the threshold. The experimental control software was written in MATLAB<sup>12</sup> using the Psychophysics Toolbox<sup>2</sup>, which provides high level access to the C-language VideoToolbox<sup>13</sup>.

The target contrast threshold was measured upon several pedestal contrasts ranged from -34dB (2%) to -6 dB (50%). On each trial, the two flankers always had the same contrast. The contrast of the flanker was either 50% (-6dB) or 0%. Each target threshold was measured at least four times for each observer. The thresholds reported here are average of those repeated measurements.

Two observers served in this study. CCC (male, early 30s) is an author of this paper. MDL (female, late 20) is a paid observer naïve to the purpose of the experiment. MDL has a normal and CCC has a corrected-to-normal visual acuity (20/20).

### 3. Results

Figure 3 shows the result from one of the observer. The smooth curves are fits of the sensitivity modulation models discussed below. The closed circles and solid curves show the TvC function measured without the presence of the flankers. When the flanker is absent, the TvC function shows the typical "dipper" shape commonly seen in the spatial contrast discrimination literature<sup>6,7,10,16,25</sup>. That is, the target threshold first decreases and then increases with pedestal contrast. The greatest threshold decrement occurs when the pedestal contrast is at about its own detection threshold. A particularly robust facilitation effect of -9 db is seen in this example.

The open circles and dashed curve show the TvC function measured in the presence of -6dB (50%). The flankers have three major effects on TvC functions. First, when there was no pedestal (denoted as  $-\infty$  dB contrast pedestal in Figure 3), the flankers reduced the target threshold.

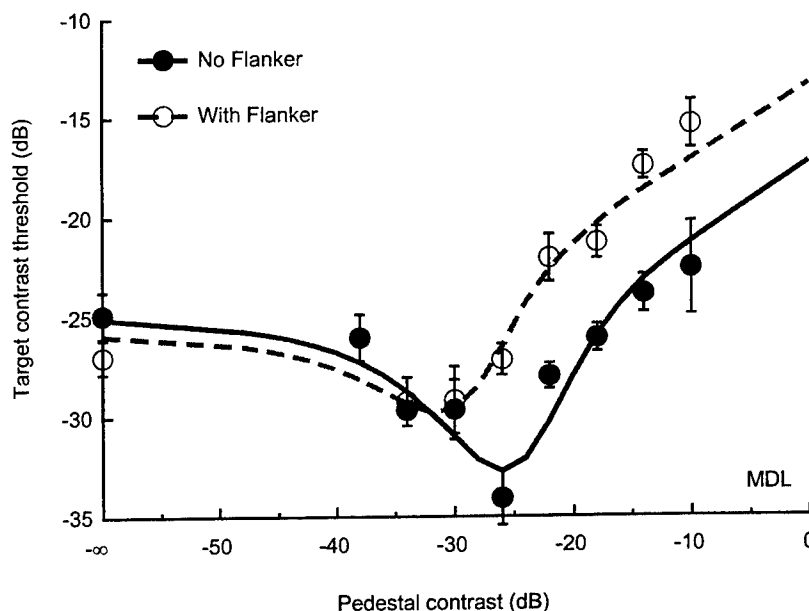


Figure 3. Flanker effect on TvC functions for observer MDL. The closed curves and solid curve denote the TvC functions measured with no flanker presented. The open circles and the dotted curve denote the TvC function measured with the presence of -6 dB or 50% contrast flankers.

This is the lateral masking effect reported by Polat & Sagi<sup>14,15</sup>. Second, as the pedestal contrast increased, the target threshold did not show as much decrement as in the case of no flankers. There was little, if any, low pedestal contrast dip when the flankers were presented. Third, at high pedestal contrasts, the flankers increased the threshold at every pedestal contrast evenly. This effect can be viewed as shifting the TvC function horizontally to the left. The two TvC functions are parallel to each other up to the highest pedestal contrast available from our apparatus. There is no sign of merging between the TvC functions measured with and without flankers presenting. The data of different observers are consistent with this result. Thus, the data reject the divisive inhibition models, which predict the two TvC functions should merge together at high pedestal contrast (see section 1.3 and Figure 2). For data shown in Figure 3, the divisive inhibition models can underestimate the target by about 6dB or 2-fold linear contrast unit.

## 4. Modeling

### 4.1 The sensitivity modulation model

In Figure 3, the data are plotted in log-log coordinates. The flankers shift the TvC functions horizontally on these coordinate, suggesting that the flanker effect is multiplicative. Based on this idea, we propose a sensitivity modulation model to account for the flanker effect.

Figure 4 shows a diagram of this model. Our model does share the same linear filter assumption as most models of contrast discrimination. That is, the visual system contains a band of localized linear filters each responds to a Fourier component of input images. Each filter has a limit extent in the space domain and bandwidth in Fourier domain. The output of each linear filter (called excitation, denoted as  $E$  in Figure 4) is the contrast of the input image weighted by a number. The weight is called the sensitivity of the filter to the image and is determined by the cross-correlation of the spatial profile of the filter and the image. A nonlinear response operator follows each linear filter. For a reason that will be obvious shortly, this direct link between the linear filter and nonlinear operator is called the excitatory input to the nonlinear operator. In addition to the excitatory input, the response of the nonlinear operator is also influenced by the other linear filters with a receptive field covering the same spatial location but tuned to different Fourier components of the input image (inside the dotted box in Figure 4). A pooling process combines the excitations of all relevant linear filters together to form the divisive inhibition signal, denoted as  $I$  in Figure 4. Mathematically, the pooling is done by summing the excitations of relevant filters raised by a power  $q$ . When the flanker is not presented, the response of the nonlinear operator is simply the excitation from the excitatory input ( $E$ ) raised by a power  $p$  and then divided by the divisive

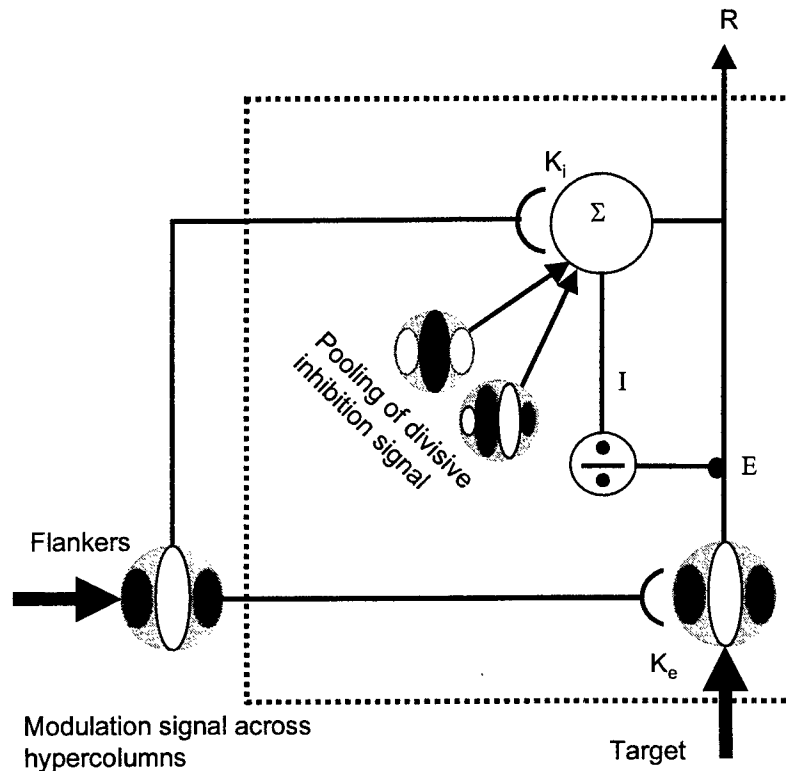


Figure 4. A diagram of the sensitivity modulation model. Inside the dotted box, all linear filters respond to image components presented at the same location. There behavior is described by the divisive inhibition models. The initial excitation ( $E$ ) of a linear filter is the contrast of the target pattern weighted by the filter's sensitivity to that pattern. The initial excitations of all relevant filters are pooled together to form the divisive inhibitory signal ( $I$ ). The final response is the initial excitation raised by a power and then divided by the normalization signal plus a constant. The flanking filters send signals that change the sensitivities of the contacted filters. See text for further details.

inhibitory input ( $I$ ) plus an additive constant. That is,

$$R = \frac{E^p}{I + \sigma} \quad (1)$$

where  $\sigma$  is an additive constant. This same-location operation is the same as that proposed by the divisive inhibition model.

In contrast to the divisive inhibition model, the flanking linear filters, however, do not contribute anything to the pooling of the divisive inhibitory signal. Instead, their effect is to change the sensitivities of the linear filters located in the same space as the target as well as the mechanism that pools the divisive inhibitory signals. That is, when the flanker is presented, the excitation of the same-location linear filters is the no-flanker excitation multiplied by a factor. After an algebraic operation, the response of the target mechanism can be written as

$$R = \frac{k_e * E^p}{k_i * I + \sigma} \quad (2)$$

where  $k_e$  and  $k_i$  are multiplicative constants dependent on the flanker contrast. When the flanker is not presented,  $k_e = k_i = 1$ . In the presence of (-6dB) 50% contrast flanker,  $k_e$  and  $k_i$  are empirically determined as 1.52~2.63 and 1.92~4.09 respectively. The fits of the model are shown as smooth curves in Figure 3. The model captures all aspects of the flanker effects. The RMSE of the model is between 0.98~1.11, on a par with the standard error of measurement (0.92~1.06).

#### 4.2 Modulation factors

How can the two factors  $k_e$  and  $k_i$  explain the flanker effects? When the pedestal contrast is low, the linear filter excitation ( $E$ ) is dominated by the target contrast. At threshold, the divisive inhibitory input to the nonlinear operator ( $I$ ) is negligible comparing to the additive constant  $\sigma$ . The response function with flankers presented can be simplified as  $R = k_e * E^p / \sigma$ . Thus, a  $k_e$  larger than 1 will boost the response and make the target easier to detected. This explains the lateral masking effect found by Polat & Sagi and the initial facilitation at lower end of the TvC functions.

As pedestal contrast increases, the divisive inhibitory inputs ( $I$ ) begin to catch up. Since  $k_i$  is larger than  $k_e$ , the flankers have a greater effect in the denominator of the response function than in the numerator. Therefore, the facilitation effect observed at low contrasts should decrease with the pedestal contrast. At medium contrasts, where the TvC function measured without the flanker presented shows a dip, the flankers produce less threshold reduction than at lower contrasts. Compared to the initial facilitation, the presence of the flankers has the effect of reducing, if not eliminating, the dip at medium contrasts. As pedestal contrast further increases, the flanker effect on the denominator of the response function eventually outweighs its effect on the numerator. The presence of the flanker is then increasing the target threshold rather than decreasing it. Finally, when the pedestal contrast is sufficiently high, the additive constant  $\sigma$  is negligible comparing to the inhibitory input  $I$ . Thus, we can simplify equation (1) as  $(E^p/I)$  and equation (2) as  $(k_e/k_i) * (E^p/I)$ . That is, the response function with the flankers presented is a constant times the response function without flankers. Translating the responses to thresholds, it gives the horizontal shift of TvC functions we observed on a log-log coordinate.

#### 5. Conclusion

We found that the presence of flankers has the following effects on TvC functions. First, when there is no pedestal, the flankers reduce the target threshold. Second, the magnitude of the dip at low to medium pedestal contrast is greatly reduced. Third, at high pedestal contrasts, the flankers shift the TvC function horizontally to left on a log-log coordinates. The two TvC functions are parallel to each other up to the highest pedestal contrasts available. Current divisive inhibition models cannot account for these effects. At some points, the discrepancy between the model and the data can be as large as 6dB or 2-fold. This discrepancy may affect the accuracy of an image quality metric. We refined the divisive inhibition models and propose a lateral sensitivity modulation model. This model takes the flanker effect as a multiplicative sensitivity modulator and captures all aspects of the flanker effect observed in our experiment.



## Acknowledgements

This study was supported by NIH grants EY7890 to CWT and a Rachel C. Atkinson Fellowship from the Smith-Kettlewell Eye Research Institute to CCC.

## References

1. Albrecht, D. G. & Geisler, W. S. (1991). Motion selectivity and the contrast response function of simple cells in the visual cortex. *Visual Neuroscience*, **7**, 531-546.
2. Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, **10**, 433-436.
3. Carandini, M. & Heeger, D. J. (1994). Summation and division by neurons in primary visual cortex. *Science*, **264**, 1333-1336.
4. Chen, C. C., Foley, J. M. & Brainard, D. H. (2000). Detection of chromoluminance patterns on chromoluminance pedestals II: model. *Vision Research*, **40**, 789-803.
5. Daley, S. (1993). The visible differences predictor: an algorithm for the assessment of image fidelity quality. In A. B. Watson (ed.). *Digital images and human vision*. MIT Press, Cambridge, MA.
6. Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, **11**, 1710-1719.
7. Foley, J. M. & Chen, C. C. (1997). Analysis of the effect of pattern adaptation on pattern pedestal effects: A two-process model. *Vision Research*, **37**, 2779-2788.
8. Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, **9**, 181-197.
9. Kontsevich, L. L. & Tyler, C. W. (1999) Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, **39**, 2729-2737.
10. Kontsevich, L. L. & Tyler, C. W. (1999). Nonlinearity of near-threshold contrast transduction. *Vision Research*, **39**, 1869-1880.
11. Legge, G.E. & Foley, J.M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, **70**, 1458-1470.
12. MathWorks (1993). *Matlab*. Natick: The MathWorks Inc.
13. Pelli, D.G. (1997). The Video Toolbox software for visual psychophysics: Transform numbers into movies. *Spatial Vision*, **10**, 437-442.
14. Polat, U. & Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research*, **33**, 993-999.
15. Polat, U. & Sagi, D. (1994). The architecture of perceptual spatial interactions. *Vision Research*, **34**, 73-78.
16. Ross, J. & Speed, H. D. (1991). Contrast adaptation and contrast masking in human vision. *Proceedings of Royal Society London, Ser. B*, **246**, 61-69.
17. Snowden, R. J. & Hammett, S. T. (1998). The effects of surround contrast on contrast thresholds, perceived contrast and contrast discrimination. *Vision Research*, **38**, 1935-1945.
18. Solomon, J. A., Watson, A. B. & Morgan, M. J. (1999). Transducer model produces facilitation from opposite-sign flanks. *Vision Research*, **39**, 987-992.
19. Teo, P. C. & Heeger, D. J. (1994). Perceptual image distortion. *SPIE proceedings*, **2179**, 127-141.
20. Tyler, C. W. & McBride, B. (1997). The Morphonome image psychophysics software and a calibrator for Macintosh systems. *Spatial Vision*, **10**, 479-484.
21. Watson, A. B. (1987). The cortex transform: Rapid computation of simulated neural images. *Computer Vision, Graphics, and Image Processing*, **39**, 311-327.
22. Watson, A. B. (1993). *Digital images and human vision*. Cambridge MA: MIT Press.
23. Watson, A. B. & Solomon, J. A. (1997). A model of visual contrast gain control and pattern masking. *Journal of the Optical Society A*, **14**, 2378 - 2390.
24. Watson, A. B., Taylor, M. & Borthwick, R. (1997). Image quality and entropy masking. *SPIE Proceedings*, **3016**, 2-12.
25. Wilson, H.R., McFarlane, D.K. & Philips, G.C. (1983). Spatial frequency tuning of orientation selectivity units estimated by oblique masking. *Vision Research*, **23**, 873-882.



## **SESSION 3**

# **Digital Camera Design and Applications**

# Measurement of the spatial frequency response (SFR) of digital still-picture cameras using a modified slanted edge method

Wei-Feng Hsu, Yun-Chiang Hsu, and Kai-Wei Chuang

40, Chungshan North Road, 3rd Sec., Taipei, Taiwan 104, ROC

Institute of Electro-Optical Engineering, Tatung University

## ABSTRACT

Spatial resolution is one of the main characteristics of electronic imaging devices such as the digital still-picture camera. It describes the capability of a device to resolve the spatial details of an image formed by the incoming optical information. The overall resolving capability is of great interest although there are various factors, contributed by camera components and signal processing algorithms, affecting the spatial resolution. The spatial frequency response (SFR), analogous to the MTF of an optical imaging system, is one of the four measurements for analysis of spatial resolution defined in ISO/FDIS 12233, and it provides a complete profile of the spatial response of digital still-picture cameras. In that document, a test chart is employed to estimate the spatial resolving capability. The calculations of SFR were conducted by using the slanted edge method in which a scene with a black-to-white or white-to-black edge tilted at a specified angle is captured. An algorithm is used to find the line spread function as well as the SFR. We will present a modified algorithm in which no prior information of the angle of the tilted black-to-white edge is needed. The tilted angle was estimated by assuming that a region around the center of the transition between black and white regions is linear. At a tilted angle of 8 degree the minimum estimation error is about 3%. The advantages of the modified slanted edge method are high accuracy, flexible use, and low cost.

**Keywords:** Digital still-picture cameras, spatial resolution, spatial frequency response, modulation transfer function, slanted edge method

## 1. INTRODUCTION

The spatial resolution capability, one of the most important attributes, of an electronic still picture camera is the ability of the camera to capture fine details found in the original scene. For electronic still picture cameras the resolving ability depends on many factors, including the performance of the optical imaging lens system, the number and the pitch of camera sensing photodetectors, as well as the electrical circuits of the functions including the gamma correction function, digital interpretation, color correction, and the image compression. There are different measurement methods which provide different metrics to quantify the resolution of an electronic camera. These metrics contain visual resolution, limiting resolution, spatial frequency response (SFR), modulation transfer function (MTF), optical transfer function (OTF), and aliasing ratio. The SFR depicts the frequency response at all spatial frequencies of a digital still-picture camera. A standard SFR algorithm employing the slanted-edge method is adopted in ISO 12233 in which a test chart containing some black-to-white and white-to-black edges, tilted at certain angles, is used to evaluate the SFR [1], [2]. In the selected region of the chart image, each row of the edge spread image is an estimate of the camera edge spread function (ESF). Each of these ESFs is differentiated to form its discrete line spread function (LSF). To accomplish this, it is first to find the position of the centroid of each row LSF which is used to find the shift of this LSF to a reference origin. It then needs to truncate the numbers of rows of data to a full cycle of rotation. The next step is the super-sampling and averaging to form a compositive requantized ESF over a discrete temporal variable which is four times more finely sampled than the original ESF. The averaged, super-sampled ESF is then differentiated and windowed to yield the LSF. The SFR is obtained using the normalized discrete Fourier transform of the single line spread function.

We have developed an algorithm to estimate the angle of a tilted edge and then to find the SFR using the curve fitting

technique by applying a mathematical model analog to the edge variation. This SFR algorithm can be applied to any test chart containing edges slanted at arbitrary angles and provide high accuracy of the SFR measurements of commercial still cameras. Without necessarily knowing the angle of a particular test chart in advance or precise alignment between the test chart and the camera, this algorithm can easily be used both in the lab and in the field.

## 2. THE SFR ALGORITHM

Figure 1 shows a flowchart of the algorithm developed for this study. The key issue of finding a precise SFR is the estimation of a correct shift of the scanning row with respect to the camera sensor grid on the chart image. The estimation of the position shift in the ISO algorithm is achieved locally by finding the difference between the closest pixel to the Centroid on each row and the Centroid. Unlike the ISO algorithm, the presented algorithm calculates the row shift from global data by finding the tilted angle between the edge and the sensor grid.

In this algorithm, after an edge area is determined, the Centroid of the area is obtained from the whole area in order to minimize the effect of random noise. The next step is to find the edge slopes on each sensor row and column (in the horizontal and vertical directions) that crosses the edge. These slopes should be found at the half of the edge height. However, the half-height slope cannot exactly be found because of the discrete nature of digital cameras. To solve this problem, those pixels with a value close to the Centroid would be used only, and the slopes are calculated from those pixels. We first set a small region, called the linear region, on each row and column around the Centroid and look for enough pixels to estimate a slope. If no enough pixels are found to find the slope, the linear region is increased until a valid number of slopes are found. In order to minimize the noise effect, the means of the row slopes and column slopes are obtained. The tilted angle  $\theta$  of the edge to the sensor row is then obtained by [3]

$$\theta = \tan^{-1} \left( \frac{\text{Mean Slope of the Columns}}{\text{Mean Slope of the Rows}} \right). \quad (1)$$

The row shift is given by

$$\Delta x = Y \cdot \tan \theta, \quad (2)$$

where  $Y$  is the pitch in the vertical axis of the camera sensor. Since the row shift is obtained, the sensor rows can be merged by properly shifting to a multiple of  $\Delta x$  to compose a highly sampled ESF. Then, the composite ESF is curve fitted with a Fermi function

$$f(x) = b + \frac{h}{1 + \exp(-w \cdot (x - c))}. \quad (3)$$

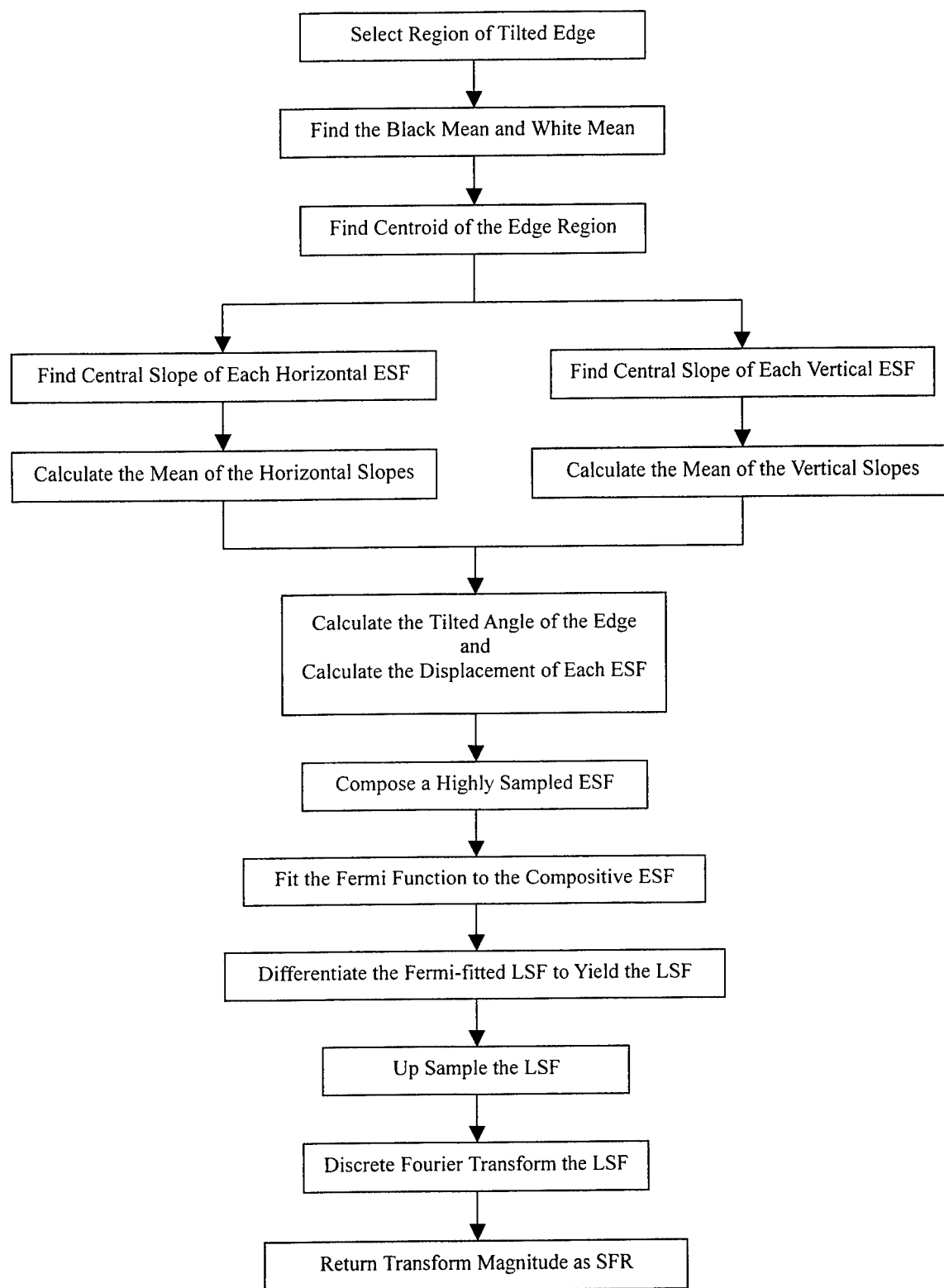
Here,  $b$  is equivalent to the mean black level on the chart image,  $h$  is the height of the ESF,  $w$  is the width parameter, and  $c$  corresponds to the center of the function. When the curve fitting is accomplished, a set of these parameters can be directly applied to the derivative of the Fermi function

$$f'(x) = -\frac{w \cdot h \cdot \exp(-w \cdot (x - c))}{[1 + \exp(-w \cdot (x - c))]^2} \quad (4)$$

which yields a continuous LSF of the edge.

Then, the curve fitting technique is employed to model the sharp of the edge transition, or the edge spread function (ESF), with the Fermi function [3], and yields a set of the parameters  $b$ ,  $h$ ,  $w$  and  $c$ . The continuous line spread function (LSF) is found by directly differentiating the obtained ESF and substituting these Fermi parameters into the differentiation of ESF. The continuous LSF is sampled by a frequency that is four times of the original sampling frequency in which the multiple of four is designated by ISO. Finally, the super-sampled LSF sequence is discrete Fourier transformed to generate the SFR of the test camera.

Input to this algorithm is a two-dimensional array containing the digital data of an image of a slanted edge. The size of this image array needs to consist of enough rows of data, typically more than 10 rows, and black and white areas, each more than 1/4 of the slanted edge image. The simulations were achieved using MATLAB programs.



**Figure 1.** Flowchart of SFR measurement algorithm

### 3. SIMULATION RESULTS

We first generated a sequence of images on which a black-to-white edge is tilted at angles of 5 to 80 degrees at an interval of 5 degrees. These edge images were sampled by assigning a set of the sensor pitches and pixel dimensions in to simulate the sampling process of a digital camera. The SFR algorithm is applied to an image of a black-to-white edge tilted at an angle ranging from  $5^\circ$  to  $20^\circ$ . Figure 2 shows the simulation of an image of the tilted edge that was generated by a computer. Each square on this image represents an area where its optical power is collected by a CCD sensor pixel. The image of the sampling result is shown in Fig. 3(a) and a composite edge-spread function of the slanted edge in Fig. 3(b) after the algorithm was applied. Here, the estimation of the angle and the selection of the function to model the edge transition are two critical issues to achieve a good approximation of the SFR. Without any noise involved, the estimation of the ESF is quite good as shown. However, various photographic situations such as different tilted angles, pixel pitches and dimensions, signal-to-noise ratios, and contrast ratio all may influence the estimation results and need to be studied in details.

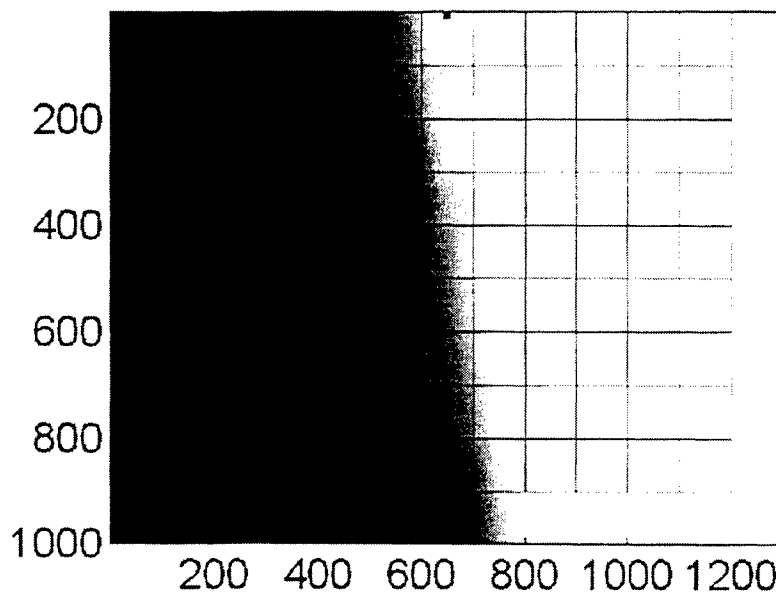


Figure 2. A computer-generated image of the tilted edge

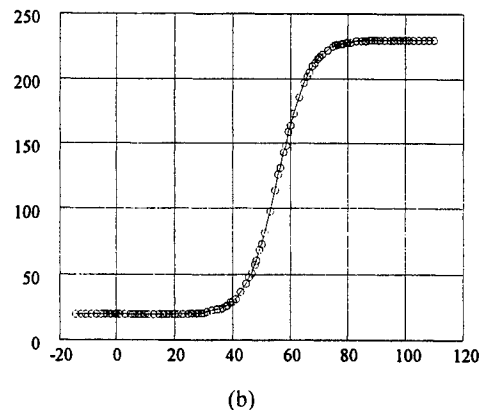
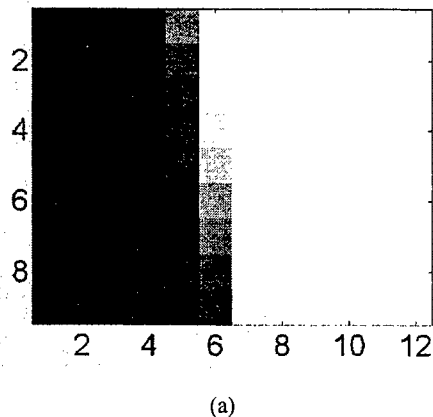


Figure 3. (a) The edge image after sampled and (b) a composite edge spread function

### 3.1 Tilted angle

The SFR algorithm was first used to find the angle of edge which is tilted from  $5^\circ$  to  $80^\circ$  in an interval of  $5^\circ$ , and the estimation results are shown as in Fig. 4. Figure 4(a) depicts the estimation angles to the given angles and their RMS errors in Fig. 4(b). The smaller RMS errors occur at small (less than  $20^\circ$ ) and large (larger than  $70^\circ$ ) angles, as well as in the middle  $45^\circ$ . Because the vertical (column) and the horizontal (row) slopes are calculated in the same way, the estimation angle should not vary significantly in the symmetric angles to  $45^\circ$ , e.g.  $10^\circ$  and  $80^\circ$ , or  $15^\circ$  and  $75^\circ$ . It is suggested according to the observation of Fig. 4 that the angles in the range of  $5^\circ$  to  $20^\circ$  provide a good estimation result to the tilted angle for this algorithm. It is noticed that the RMS error at the tilted angle  $45^\circ$  is also small. Nevertheless, it is not preferred here for the reasons discussed later.

### 3.2 Pixel pitch and dimension

In the simulation, the width of the edge transition is designed to be  $46\text{ }\mu\text{m}$  for the digital level varying from 1% to 99% of the edge height. The variables  $W$ ,  $D$ , and  $d$  denote the width of the edge, the pixel pitch, and the pixel dimension, respectively. The estimation results of three tilted angles ( $10^\circ$ ,  $30^\circ$ , and  $45^\circ$ ) are shown in Fig. 5. The normalized sampling period is defined as the ratio of the pixel pitch to the edge width, i.e.,  $D/W$ . In Fig. 5(a), the RMS error increases as the normalized sampling period increases. The errors of the edge tilted at  $45^\circ$  vary greatly at  $D/W \approx 0.5$ . A tilt of  $45^\circ$  results in a shift of a half of the pixel pitch and thus only a sampled pixel locates in the edge transition region. The poor sampling process occurs both at the vertical and horizontal directions and results in large RMS errors. It is one of the reasons that  $45^\circ$  tilted angle is not preferred.

Figure 5(b) shows the RMS error of the estimations for various aspect ratios, defined as the ratio of the pixel dimension ( $d$ ) to the pixel pitch ( $D$ ). The RMS error slightly decreases as the aspect ratio increases for the tilted angles of  $30^\circ$  and  $45^\circ$ , but remains almost constant for the angle  $10^\circ$ . The aspect ratio does not significantly affect the estimation results for the use of this algorithm.

### 3.3 Signal-to-noise ratio

It would be important and practical to analyze the performance of the presented algorithm when it is applied to an image containing noises. The RMS error versus the signal-to-noise ratio (SNR) is shown in Fig. 6. It is observed that the RMS error does not change significantly even the SNR is as low as 5 for the tilted angles of  $10^\circ$  and  $45^\circ$ , and it only roughly decreases as SNR increases for the tilted angle  $30^\circ$ . This algorithm is immune to the noise effects due to the use of the Fermi function that eliminates the noise variations at the step of curve fitting. Therefore, it is suggested that smaller tilted angles around  $10^\circ$  would be preferred in this algorithm.

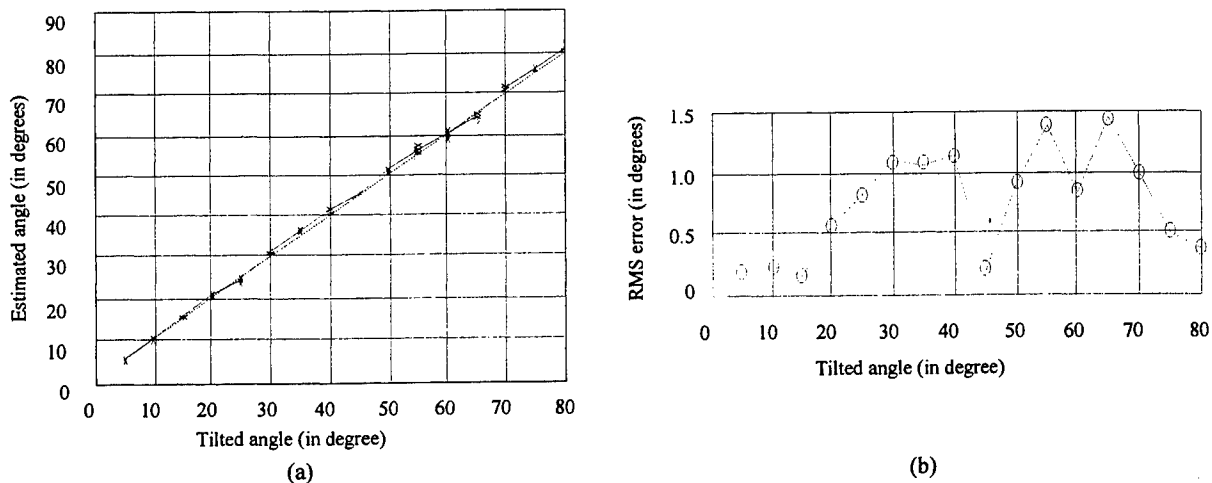


Figure 4. (a) Estimations of the tilted angles and (b) the RMS errors

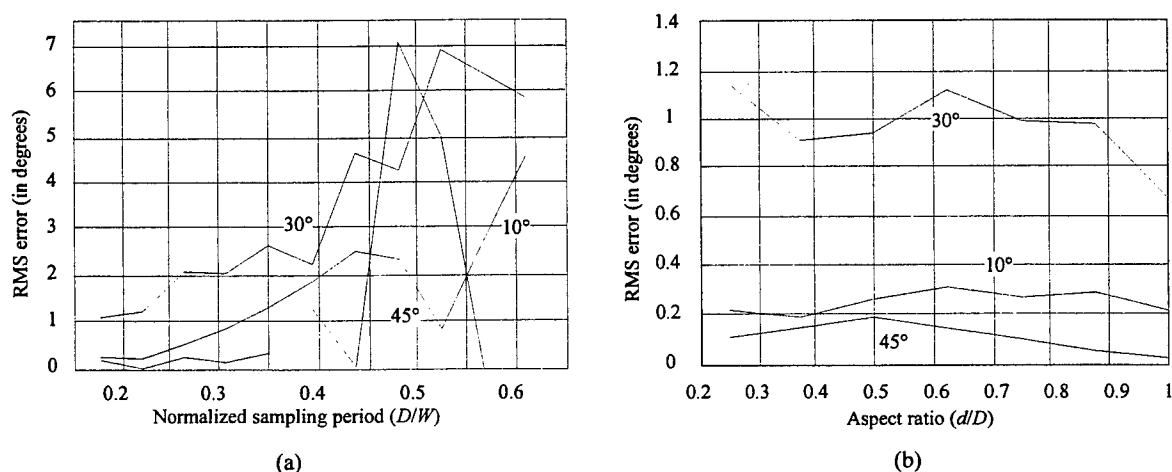


### 3.4 Contrast ratio

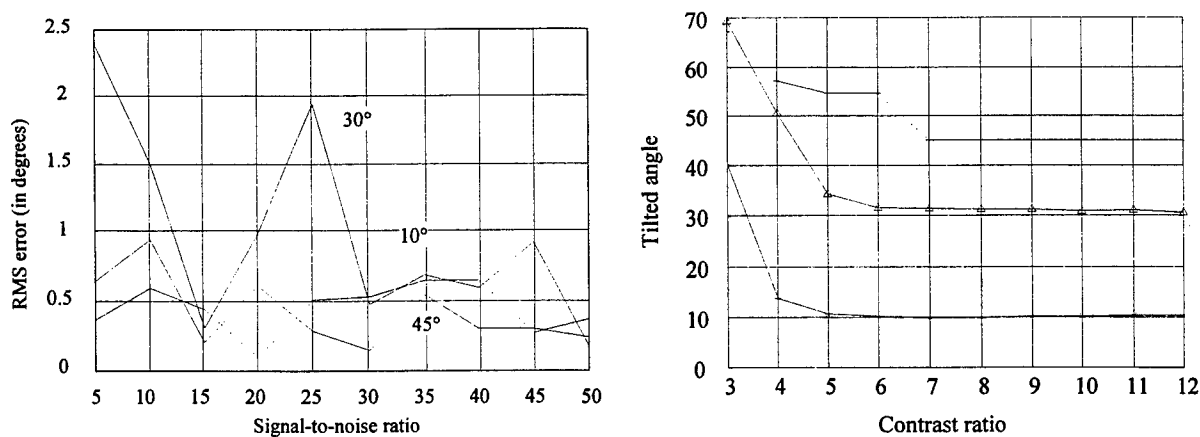
The contrast ratio is defined as the ratio of the brightness of the white area to that of the black area. As shown in Fig. 7, the estimated angle approaches to the real tilted angle for the contrast ratio greater than a value depending on the tilted angle. The value decreases as the tilted angle decreases. The edge of a tilted angle of  $10^\circ$  in an image of a contrast ratio as low as 5 can be precisely estimated using this algorithm.

### 3.5 Estimation of the spatial frequency response (SFR)

The estimation of the spatial frequency response of the edge image is shown in Fig. 8 in which the dashed line denotes the SFR of a perfect edge. In the test images, the edge is tilted at  $10^\circ$  and the SNR is given from 5 to 20. The estimation error is the difference between the estimated SFR and the perfect SFR at the modulation of 0.05. The spatial frequency at



**Figure 5.** (a) The RMS error versus the normalized sampling period (at a fixed aspect ratio of 1) and (b) the RMS error for different aspect ratio (at normalized sampling period 0.17,  $D = 8 \mu\text{m}$ )

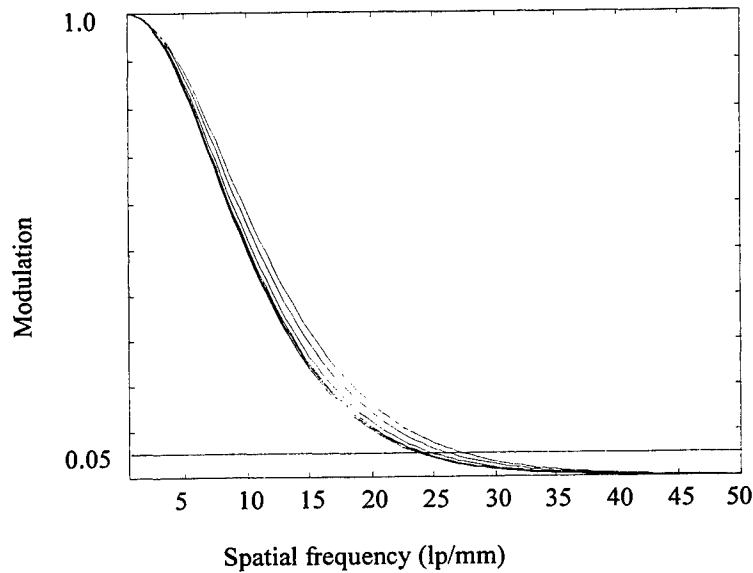


**Figure 6.** The RMS errors for different signal-to-noise ratio (at the fixed aspect ratio 1 and the normalized sampling period 0.17)

**Figure 7.** The estimation of the tilted angles for different contrast ratio (at the fixed aspect ratio 1 and the normalized sampling period 0.17)

**Table 1.** Estimation of the SFR of the edge tilted at  $10^\circ$

SFR	Standard	SNR = $\infty$	SNR = 5	SNR = 10	SNR = 15	SNR = 20
Frequency at the modulation 0.5 (lp/mm)	26.3	25	27.4	23.9	24.1	24.2
Frequency error (lp/mm)	--	1.3	1.1	2.4	2.2	2.1



**Figure 8.** Estimation of the SFR for images of SNR = 5, 10, 15, 20, and  $\infty$

the modulation 5% is used as the reference because the limiting resolution, one of the resolution metrics [4], is defined as the spatial frequency at a modulation of 0.5. It is noticed that all the frequency errors are less than 2.5 line-pairs per millimeters (lp/mm) as listed in Table 1. Note that the pixel pitch is  $10\ \mu\text{m}$  and thus the Nyquist frequency is 50 lp/mm

#### 4. CONCLUSIONS

The presented algorithm can be applied under various measurement environments since the angle information is not required for the estimation of the camera SFR and, therefore, no official test chart is needed. According to the simulations of the algorithm, it is suggested that the angle should be tilted between  $5^\circ$  through  $20^\circ$ . Although, the best estimation result occurs at the angle tilted at  $45^\circ$ , the edge of tilted angle  $45^\circ$  is not preferred because the estimation of the  $45^\circ$  angle cannot provide a stable estimation at normalized sampling periods around 0.5 and when noise happens to corrupt the single sampled pixel in the edge region. In conclusions, the advantages of the proposed algorithm are:

1. It can be used in low signal-to-noise ratio.
2. It can be used in low contrast ratio.
3. The cost of the test chart is low.

## ACKNOWLEDGMENTS

This work was supported in part by Tatung University, Taipei, Taiwan, R.O.C. under the grant B87-1011-01.

## REFERENCES

1. D. Williams, "Benchmarking of the ISO 12233 slanted-edge spatial frequency response plug-in," *IS&T's 1998 PICS Conference*, pp.133-136.
2. Sheng-Yuan Lin, Wen-Hsin Chan, Wei-Feng Hsu and Tim Y. Tsai, "Resolution characterization for digital still cameras," *IEEE Trans. Consumer Electronics*, Vol. 43, No. 3, August 1997, pp. 732-736.
3. Wei-Feng Hsu, et al., Technical Report in Opto- Electronics & Systems Lab, Industrial Technology Research Institute, July 1998.
4. ISO/DIS 12232: Photography- Electronic still picture cameras- Determination of ISO speed, 1997.

# Comparisons of the camera OECF, the ISO speed, and the SFR of digital still-picture cameras

Wei-Feng Hsu, Kai-Wei Chuang, and Yun-Chiang Hsu

40, Chungshan North Road, 3rd Sec., Taipei, Taiwan 104, ROC

Institute of Electro-Optical Engineering, Tatung University

## ABSTRACT

In this paper, the techniques as well as the measurement results of the performance of commercial digital still-picture cameras are presented. The key parameters such as the camera Opto-Electronic Conversion Function (OECF), the noise based ISO speed, and the spatial frequency response (SFR) are reported. The camera OECF is defined as the relationship between the input luminance and the grayscale or digital output from the camera, which was measured by using a test chart with twelve squares of various luminances. The ISO speed was calculated from the exposure time, the effective f-number, and the luminance at different incremental signal-to-noise ratios. In general, the exposure time is not obtainable from a commercial digital camera unless a destructive measurement is undergoing. In this study, a device was setup to obtain the exposure time when the OECF test chart was recording. A modified slanted-edge method was employed to estimate the SFR by imaging a pattern with a black-to-white edge tilted at an arbitrary angle. There are seven digital still picture cameras as our test samples whose CCD sensor contains VGA-size and million pixels. The camera OECF of these cameras did not show significant difference under a large range of illumination. However, the ISO speed and the SFR were of great variation.

**Keywords:** Digital still-picture camera, opto-electronic conversion function, incremental signal-to-noise ratio, ISO speed, spatial frequency response

## 1. INTRODUCTION

Digital still-picture cameras (DSC) using CCD or silicon CMOS area sensors have becoming a major product on the consumer electronic market. In a complex DSC system, the quality of acquired images is mainly affected by (1) the optics, such as the lens system for imaging and zooming, (2) the opto-electronic (OE) devices such as CCD area sensors, and (3) analog and digital electrical circuitry. In addition, various signal-processing algorithms such as data interpretation, color correction, image compression, etc. also influence the image quality in some way. For DSC manufacturers, it is important to understand and exactly characterize the camera performance in order to improve the image quality and design better, convenient usages to the cameras. For users, a simple and clear illustration to the camera will assist to purchase a proper camera to their needs.

However, it is difficult to provide a complete, satisfactory analysis of the performance of electronic still-picture camera because it is constructed by most part of the conventional film imaging cameras and an electronic microprocessor. ISO and many research groups have been working on standardizing the characteristics and test methods of digital still-picture cameras. The general characteristics of interest include the relationship of the optical input to the digital output level, the noise performance, the equivalent exposure speed, the spatial resolving capability, and the color performance. Each again cannot be fully described by a single function or parameter using a single method. In the ISO documents, some functions and various methods are defined for each characteristic in order to fit different camera organizations and test environments. Complete measurements of these characteristics would be tedious and unnecessary. In this study, we select an opto-electronic conversion function (OECF) method, an ISO speed, and the spatial frequency response (SFR) in the resolution measurement to evaluate the performance of commercial electronic still-picture cameras. There are seven cameras used for in this study: three have a million-pixel CCD sensor, Case 1 to Case 3, and four have a VGA-size CCD sensor, Case 4 to Case 7. Limited to the source of digital still-picture cameras, this research does not intend to provide a

performance benchmark. Our goal is to develop effective and simple methods to precisely characterize the digital still-picture cameras, to improve their performance, and to develop various functionalities of digital cameras based on the techniques presented in this report. Some other works can be referred at the references [1]-[3].

## **2. CHARACTERISTICS OF DIGITAL STILL-PICTURE CAMERAS**

### **2.1 Opto-Electronic Conversion Functions (OECF's)**

In ISO14524 [4], the opto-electronic conversion function (OECF) is defined as the functional relationship between the optical input and the digital output signal level of a digital still-picture camera. The measurement of the OECF's is fundamental because it is required for the development of digital cameras and for the calculations of other characteristics such as the ISO speeds and the spatial resolution. Besides, it will be used for data correction for other digital still-picture camera characteristics and may be helpful in the processes of digital image data.

Although analog terminologies, such as the "H&D" curve for photographic films and the "gamma" curve for CRT monitors, are widely used, the OECF's are necessary because none of these methods can be easily or unambiguously applied to electronic still picture cameras. In digital systems, the sampling and quantization processes present fundamental issues that need to be addressed in a standardized manner. The flexibility of digital systems complicates the determination and presentation of the functional relationship between the camera's optical input and digital output signal level. Therefore, ISO 14524 attempts to account for all the variables and assure that results are presented in consistent fashion.

There are three OECF's defined in ISO 14524. The first is the camera OECF that is accomplished by using the camera system to capture an image of the chart under controlled conditions. Then, the focal plane OECF involves the exposure of the electronic still picture camera sensor directly to specific quantities of uniform illumination with the camera lens removed. The alternative focal plane OECF is introduced for the measurement of focal plane OECF under the circumstances that a particular electronic still picture camera does not allow the lens to be removed. It is suggested that the focal plane OECF to be reported mainly because the focal plane OECF provides the optical-input-to-digital-output relationship of the camera part including the sensor and the signal processing electronics. It behaves quite differently from the optical image formation part which converts scene into image and is scene dependent. This tends to be highly non-linear and complicates further analysis. We therefore prefer to analyze these two parts separately. However, the mandatory automatic exposure control and the fixed imaging lens system found in lots of digital cameras preclude the determination of focal plane OECF.

### **2.2 ISO speeds**

The ISO speed rating is an important reference of photographic systems in order to guide the users to set a proper exposure when taking a photograph. Many factors affect the camera exposure including the exposure time, the iris aperture, the lens transmittance, the illumination, and the scene reflectance [5]. Using an insufficient exposure to obtain an image could result in a noisy image for the electronic gain being applied automatically. On the other hand, if the exposure exceeds a certain amount, says the saturation of camera sensors, the image becomes bloom in the bright areas and blur on the edges. The contrast is also decreased. With the help of the ISO speed ratings, the user can set proper exposures and the manufacturers can provide more functional settings to the exposure control of the cameras.

An ISO speed rating is intended to serve as a guide to photography. The ISO speed rating for an electronic camera should directly relate to the ISO speed rating for photographic film cameras. For example, if an electronic camera has an ISO speed of ISO 100, then the same exposure time and aperture be appropriate for ISO rated film / process system. However, differences exist between electronic and conventional film cameras. Electronic cameras have a range of ISO speed ratings because the best quality of image can be achieved by varying the electronic gain and by performing signal processing algorithms after the image is captured. In ISO 12232, this range is defined as the ISO speed latitude.

Two types of the ISO speed are defined: Saturation based speed and noise based speed. The later speed is preferably used to indicate the camera's underexposure latitude, and the former speed value is preferably used to the camera's overexposure latitude. Since the camera view depth is enhanced when the exposure is lower, the noise based value is preferably determined for most electronic cameras. In this report, the noise based ISO speeds of several electronic cameras were presented.

## 2.3 Resolution measurement

The spatial resolution metrics, the test methods, and the test charts were standardized in ISO 12233 [6]. The spatial resolution capability, one of the most important attributes, of an electronic still picture camera is the ability of the camera to capture fine details found in the original scene. For electronic still picture cameras the resolving ability depends on many factors, including the performance of the optical imaging lens system, the number and the pitch of camera sensing photodetectors, as well as the electrical circuits. The functions of the electrical circuits in camera include the gamma correction function, digital interpretation, color correction, and the image compression.

Different measurement methods can provide different metrics to quantify the resolution of an electronic camera. These metrics contain visual resolution, limiting resolution, spatial frequency response (SFR), modulation transfer function (MTF), optical transfer function (OTF), and aliasing ratio. In ISO 12233, a resolution test chart is designed to evaluate the resolving performance of an electronic camera. The visual resolution is subjectively judged based on vertical, horizontal and diagonal hyperbolic wedges on the test chart. These patterns should be enlarged by integer multiples on a hard copy and then reproduced on a monitor or a printer to ensure that the measurement value is not reduced due to the different resolution between the test camera and the output device. The limiting resolution is determined by calculations of the resolution response of vertical and horizontal square wave sweeps on the test chart. The spatial frequencies of the square wave sweeps vary from 1/100 to 1/1000 of the height of the test chart. The value (in the unit of line-widths-per-picture-height, or LW/PH) is found at the location where the resolution response of the square wave is equal to 5% of the reference response. The two resolution metrics described above can only obtain discrete measurements on certain spatial frequencies provided by the test chart. The SFR measurement provides an overall frequency response of an electronic camera by capturing an image of a test pattern that contains a square tilted at a small angle. A computer then analyzes the image data, consisting of a tilted black-to-white edge, by using a standard SFR algorithm.

In this report, the spatial frequency response (SFR) was evaluated by using an algorithm that applies the same concepts as the ISO standard SFR algorithm. The algorithm adopted here can be applied to any test pattern containing black-to-white (and white-to-black) edge tilted at arbitrary angles, although it is optimized for the tilt angle between 10 to 25 degrees. Moreover, the method has high reliability even for an image of a low S/N ratio such as 5. Therefore, it is preferably adopted at a laboratory or a production line where precise alignment equipments cannot be used.

## 3. CHARACTERISTICS MEASUREMENTS OF ELECTRONIC STILL-PICTURE CAMERAS

### 3.1 Measurements of the camera OECF

In this study, the camera OECF's of some commercial electronic cameras were measured. The camera OECF is selected because of not only the reasons discussed in Sect. 2.1 but also some practical issues. The measurement of the camera OECF includes the effects of the camera lens and associated flare, while focal plane OECF's do not. With the image formation effects vary with the overall scene luminance ratio, this variability can quite large, and consequently it is possible to determine a repeatable camera OECF only for a specific scene, such as a test chart. Most digital still picture cameras do not allow the lens to be removed. The mandatory automatic exposure control found in some cameras precludes the determination of Focal Plane OECF's. Since the optical image formation stage is not removed, a test chart recording enough scene variations is imaged onto the camera sensor and the camera OECF is then obtained by analyzing the test-chart image. The whole processes are accomplished in one exposure, unlike many exposures are needed when the focal plane OECF is measured. The sensor illuminance shall be assumed to be as calculated from the following equation:

$$E_s = \frac{0.65 \times L_t}{A^2}, \quad (1)$$

where  $E_s$  is the illuminance in lux falling on the sensor,  $L_t$  is the arithmetic mean luminance of the target in candelas per square meter, and  $A$  is the effective f-number of the lens. Figure 1 shows a Camera OECF Test Chart, similar to the Camera OECF Test Chart designed by ISO, simulating the image formation effects produced by a scene with a specific luminance ratio and average distribution of luminances. In the test chart, there are twelve squares with neutral reflectivities varying from low to high, and each square represents a luminance value measured directly using a photometer. A thin stick driving by a stepping motor in the middle of the chart is used to estimate the exposure time. This chart is also used to determine the incremental signal-to-noise ratio and the ISO noise based speed discussed in the next section.

To determine the camera OECF, images of the test chart were recorded for computer to proceed calculations. At each test,

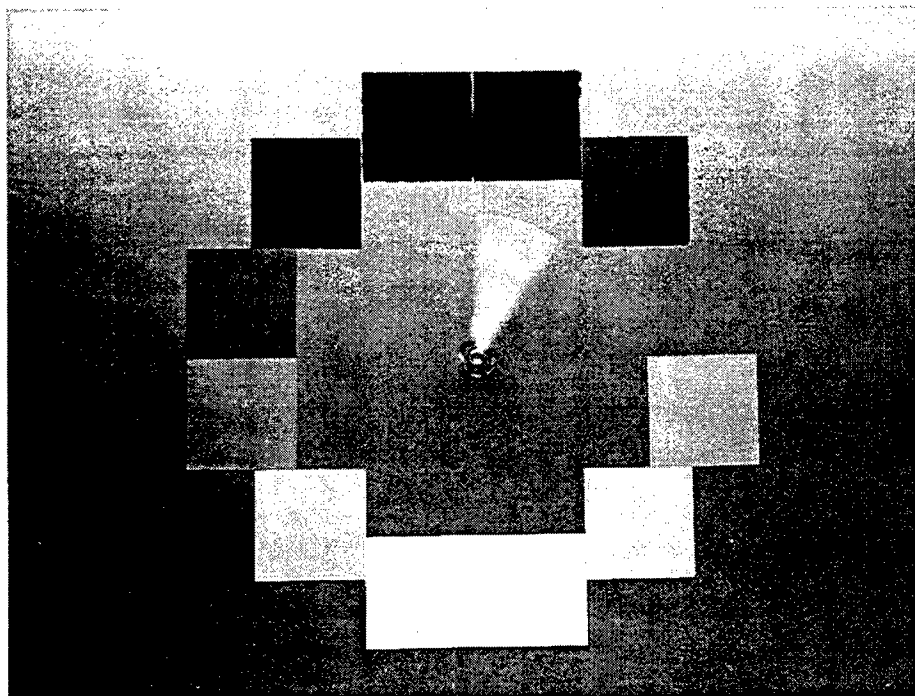


Figure 1. The image of the Camera OECF Test Chart

ten trials were taken in order to minimize random noise appearing at each exposure. For each trial, the mean digital output level was determined from a 64 by 64 pixel area located at the same relative position in each image. The 64 by 64 pixel area shall be located at the center of the test block. The final digital output level data presented is the mean digital output levels for all the trials. The input luminance was measured using PR-650 SpectraColorimeter.

The measurement of the camera OECF's of the seven test cameras are shown in Fig. 2(a). Limited to the light source in our lab, most cameras did reach their saturation except Case 5. Notice that the automatic electronic gain increases the output digital level as well as noise under the underexposure condition. Hence the digital levels of the dark areas are increased significantly. When the saturation is reached as Case 6, the dark area does look 'dark' because the electronic

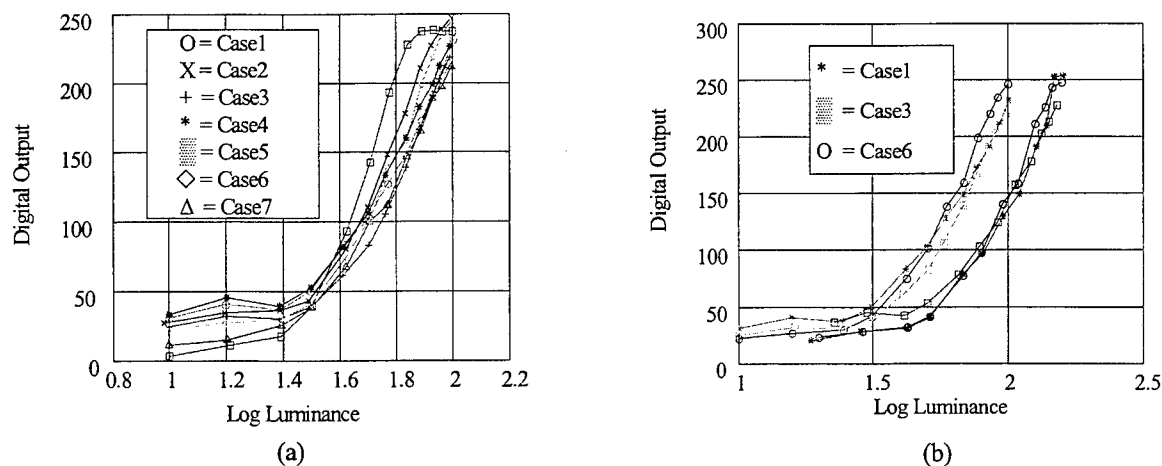


Figure 2. (a) The camera OECF of the digital still-picture cameras and (b) the camera OECF's measured at different illuminations

gain did not enlarge the output level as in the other cases. However, the saturation of Case 6 only provides an output level of 240, far below the general saturation output at 255. Figure 2(b) shows two sets of the camera OECF's measured at two illuminances in which the largest luminances from the bright square are about 160 and 100  $\text{cd/m}^2$ , respectively. The measured electronic gains are larger than 1 in this figure although the optical input was increased by 80%. This phenomenon could be improved by incorporation a contrast enhancement method with the auto-exposure process in the signal processing unit.

### 3.2 Measurements of the noise based ISO speed

In many photographic applications, it is desirable to use the highest exposure index (lowest exposure) possible, in order to maximize the depth of field, minimize the exposure time, and offer the maximum acceptable latitude for exposure of image highlights. The noise based ISO speed serves as the exposure index that provides an appropriately low noise image for typical electronic camera applications. Two different noise based speeds are determined, one that provides the "first excellent" image and a second that provides that provides the "first acceptable" image.

The noise based speed,  $S_{noiseX}$ , can be obtained using either the focal plane method or the scene luminance method. The later is adopted in this study. In this method, the ISO noise based speeds are determined from the scene luminance required to produce specific image incremental signal-to-noise ratio values using the following equation:

$$S_{noiseX} = 15.4 \times \frac{A^2}{L_{S/N} \cdot t_s}, \quad (2)$$

where  $A$  is the effective f-number of the taking lens,  $t_s$  is the photosite integration time, and  $L_{S/N}$  is the luminance that provides a camera signal-to-noise ratio,  $S/N$ , satisfying the following criterion:

$$S/N = \frac{L_{S/N} \times g(L_{S/N})}{\sigma(D_L)}. \quad (3)$$

Here,  $S/N$  is the signal-to-noise ratio of the value  $X$ ,  $L_{S/N}$  the luminance in  $\text{cd-m}^{-2}$ ,  $g(L_{S/N})$  the incremental gain (the rate of change in the mean output level divided by the rate of change in the input luminance), and  $\sigma(D_L)$  the standard deviation of the monochrome output level values (for monochrome cameras) taken from a 64-by-64 pixel area. Note that  $S_{noise42}$  denotes the noise based speed measured at  $S/N$  equal to 42, and is designated to the "first excellent" image quality;  $S_{noise10}$  denotes the noise speed measured at  $S/N = 10$ , and is designated to the "first acceptable" image quality.

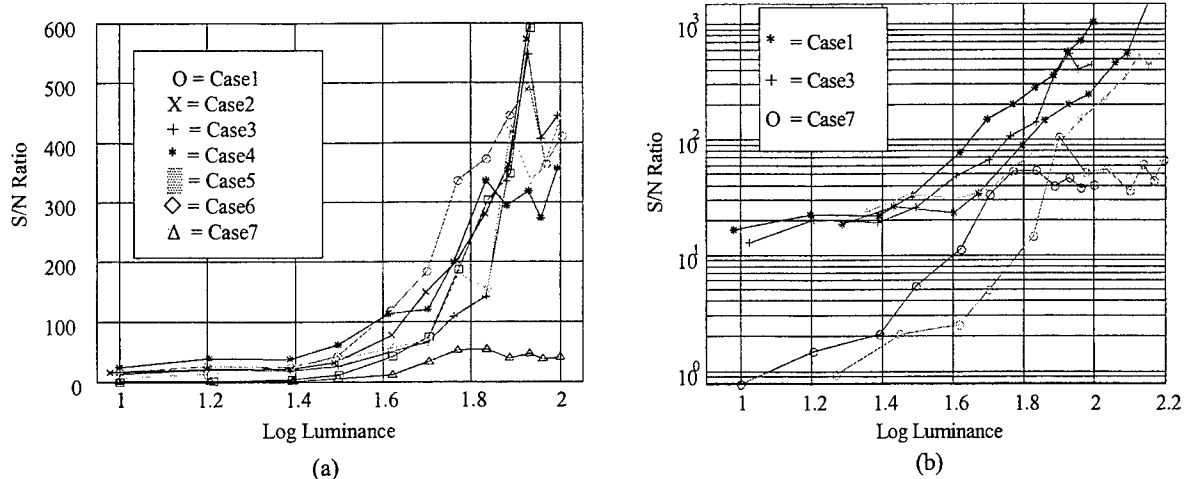
The value of  $L_{S/N}$  was determined by plotting the incremental  $S/N$  as a function of the luminance  $L$  and estimating the value that produces an incremental  $S/N$  value equal to 42 for  $S_{noise42}$  and 10 for  $S_{noise10}$ . The plot of incremental  $S/N$  to  $L$  was obtained using the same test chart as used in the measurement of the camera OECF.  $\sigma(D_L)$  was determined as the standard deviation of the pixel values in each 64-by-64 area selected for the camera OECF. The incremental gain  $g(L_{S/N})$  was also obtained from the camera OECF by using the equation

$$g(L_j) = \frac{OL(L_j) - OL(L_j - \Delta L_{i,j})}{2\Delta L_{i,j}} + \frac{OL(L_j + \Delta L_{j,k}) - OL(L_j)}{2\Delta L_{j,k}}, \quad (4)$$

**Table 1.** Noise based ISO speed of the digital still-picture cameras

Case No.	1	2	3	4	5	6	7
$S_{noise10}$	320	640	800	400	160	640	320
$S_{noise42}$	100	50	64	64	100	250	160





**Figure 3.** (a) The incremental signal-to-noise ratio of the digital still-picture cameras and (b) three incremental signal-to-noise ratio measurements at different illuminations ( $L_1=100 \text{ cd/m}^2$  and  $L_2=160 \text{ cd/m}^2$ )

where  $OL$  and  $\Delta L_{ij}$  are the digital output level and the change in luminance between luminance  $L_i$  and luminance  $L_j$ , respectively.

The noise based ISO speed was calculated by Eq. (2) in which the integration time was estimated using the clock-like timing stick in the Camera OECF Test Chart as shown in Fig. 1. The stepping motor was driving by a clock period of 1.4 ms and it took 200 steps for a cycle ( $360^\circ$ ). The stick rotated about  $30^\circ$  during the exposure which is equivalent to an integration time of 23.3 ms. The noise based ISO speed obtained from the measured values are listed in Table 1. Note that the speed value in Table 1 is determined by referring to Table 1 of ISO 12232 in which the results from Eq. (2) were assigned to a specific ISO rating for a certain range of the calculated values. For example, if an  $S_{noiseX} = 116.5$  is calculated, then the speed rating of ISO100 D should be reported for this  $S_{noiseX}$  being the range of 100 through 125. The letter 'D' denotes that the measurement was taken using a daylight illumination. Figure 3 shows the incremental signal-to-noise ratio of the test cameras. The incremental  $S/N$  is increased as the input luminance increased as expected. Because there are many factors to influence the incremental  $S/N$  ratio, it is impossible to provide a clear analysis unless the effects of individual processes in a digital camera can be measured. However, the extremely large  $S/N$  values observed in Fig. 3(b) are referred to the camera blooming phenomenon due to the sensor saturation.

### 3.3 Measurements of the SFR

Spatial frequency response (SFR) is one of the most important characteristics of digital still cameras, which describes the capability to resolve the spatial details of images generated from incoming optical information. The algorithm adopted first estimates the angle of a tilted edge, computes the edge spread function (ESF) using the curve fitting technique, gives the line spread function (LSF) by differentiating the ESF, and finally generates the SFR by Fourier transforming the LSF. The advantage of this algorithm is that it can be applied to any test chart containing edges slanted at arbitrary angles and provide high accuracy of the SFR measurements of commercial still-picture cameras. Figure 4 shows the SFR Test Chart used in this study that contains nine chessboard-like patterns and each consists of four squares, two in black arranged in diagonal and two in black arranged in back diagonal. This arrangement is designed to provide horizontal as well as vertical edges changing both from back to white and white to black in a small area on the chart for measuring the ECF of an imaging camera. Without necessarily knowing the tilted angle of a particular test chart in advance or using precise alignment between the test chart and the camera, this algorithm can easily be used both in the lab and in the field. The details of the developed algorithm and its performance analysis are presented in another paper presented in this conference.

The SFR's of the digital cameras are shown in Fig. 5. The results are consistent with the theory that the smaller the pitch between the sensor pixels, the larger the cutoff spatial frequency. The cameras of Case 1 and 3, both having a  $1/3''$  million-pixel CCD sensor, have the smallest pixel pitch (about 6 microns) and have the largest cutoff frequencies.

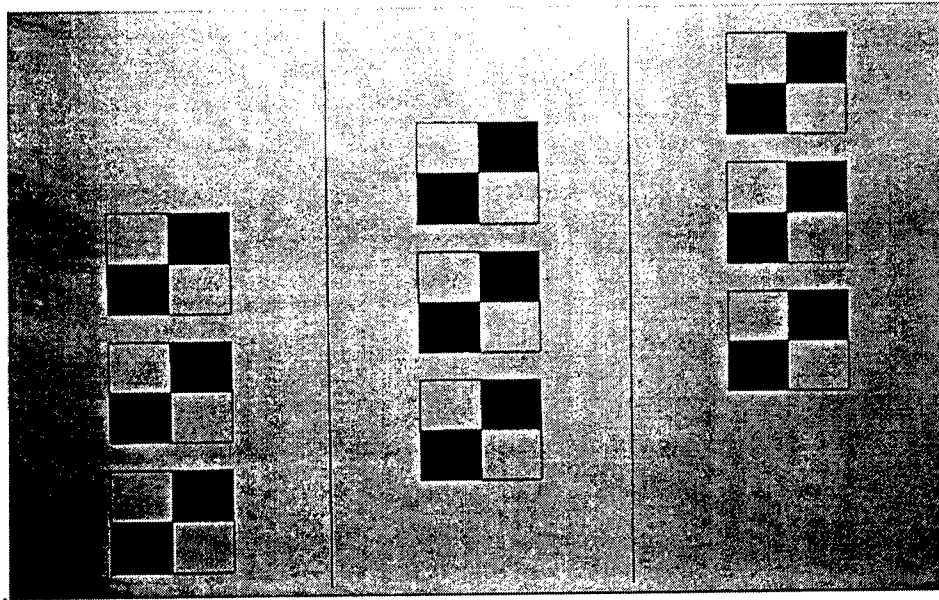


Figure 4. The image of the SFR Test Chart

Although there are many possibilities that Case 1 might have a better SFR than that of Case 3, one thing is for sure that the optics of Case 1 is more delicate. Case 2 is also a million-pixel digital camera but with a 2/3" CCD sensor and large pixel pitches. It has a poor cutoff but the SFR is almost as good as that of Case 3, and this might be due to the delicate optics on Case 2. Moreover, the SFR of Case 2 is better than those of all the VGA-size cameras. The reason that Case 5 has the worst SFR is because at the moment we used the Case-5 camera to photograph the SFR Test Chart, the condition of the camera became worse than it first came to our lab. The images were barely clear and its electronics did not function well.

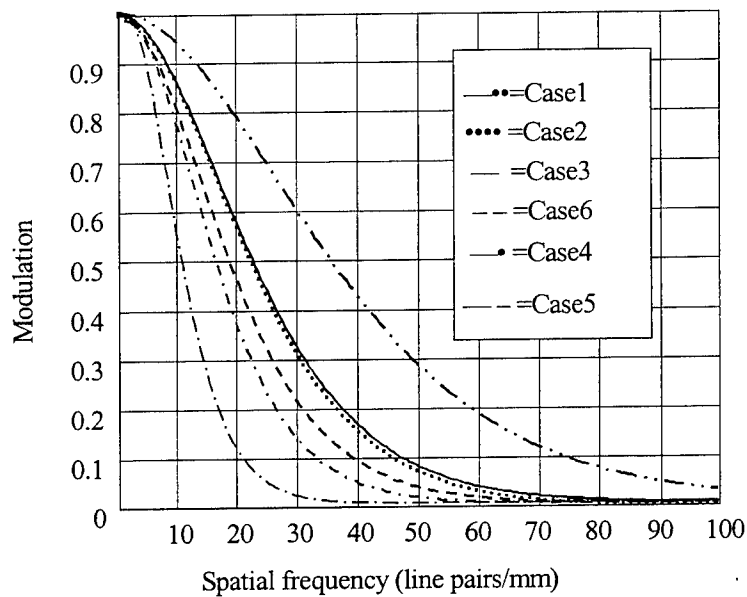


Figure 5. Spatial frequency response of the digital still-picture cameras

#### 4. CONCLUSIONS

We have presented the measurement results of the camera OECF, the incremental signal-to-noise ratio, the noise based ISO speed, and the SFR of seven digital still-picture cameras. There is no significant difference in the measured OECF's, the incremental signal-to-noise ratios, and the noise based speeds regarding to different components and structures of these cameras. It might be due to the electronics for signal processes greatly smooth the variations from different components. The influence of the electronics needs further researches and is not known for this moment. The characteristics that can be used to distinguish the performance of a digital camera will the SFR. However, the units used in the SFR, i.e., line pairs per mm, will mislead user. A more proper unit that is easily understood will be necessary in the future.

#### ACKNOWLEDGMENTS

The authors have the pleasure to thank Dr. Y. Tim Tsai at Mustek Systems Inc., Dr. Hwang-Cheng Chiang at ITRI, and other people who kindly lent their digital still-picture cameras for the measurements in this study. This work was supported in part by Tatung University, Taipei, Taiwan, R.O.C. under grant B88-2500-03.

#### REFERENCES

1. S. E. Reichenbach, S. K. Park, and R. Narayanswamy, "Characterizing digital image acquisition devices," *Optical Engineering*, Vol. 30, No. 2, pp. 170-177, February 1991.
2. A. P. Tzannes and J. M. Mooney, "Measurement of the modulation transfer function of infrared cameras," *Optical Engineering*, Vol. 34, No. 6, pp. 1808-1817, June 1995.
3. D. Williams, "Benchmarking of the ISO 12233 slanted-edge spatial frequency response plug-in," in *Proceeding of IS&T's 1998 PICS Conference*, pp. 133-136.
4. ISO/DIS 14524: Photography- Electronic still picture cameras- Methods of measuring opto-electronic conversion functions (OECF's), 1997.
5. ISO/DIS 12232: Photography- Electronic still picture cameras- Determination of ISO speed, 1997.
6. ISO/DIS 12233: Photography- Electronic still picture cameras- Resolution measurements, 1997.

# A Digital Camcorder Image Stabilizer Based on Gray Coded Bit-plane Block Matching

Yeou-Min Yeh, Sheng-Jyh Wang and Huang-Cheng Chiang

Institute of Electronics Engineering, National Chiao Tung University  
Hsinchu, 30010, Taiwan, R.O.C

and  
Industrial Technology Research Institute  
Hsinchu, 30010, Taiwan, R.O.C

## ABSTRACT

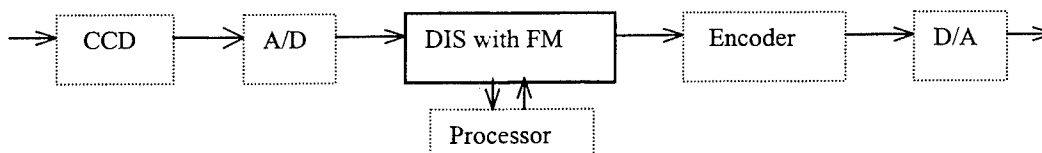
In this paper, we proposed an efficient algorithm to do image stabilization for digital camcorder. This approach is based on gray-coded bit-plane block matching to eliminate the displeasing effect caused by involuntary movement of camera holders. To improve moving object detection and stabilization performance, a frame is divided into several blocks to do localized motion estimation. Based on our architecture, the temporal correlation is used at the motion unit to efficiently detect moving objects and panning conditions. To compensate for camera rotation, an energy minimization is also applied to calculate the coefficients of affine transform without many complicated computations. Having considered both programming flexibility and hardware efficiency, the motion decision and motion compensation units are coded in a microprocessor that interconnects with the stabilization hardware. The proposed stabilizer is now implemented on FPGA 10K 100.

**Keywords:** Digital image stabilization, Motion estimation, Digital camcorder, Gray-coded bit-plane

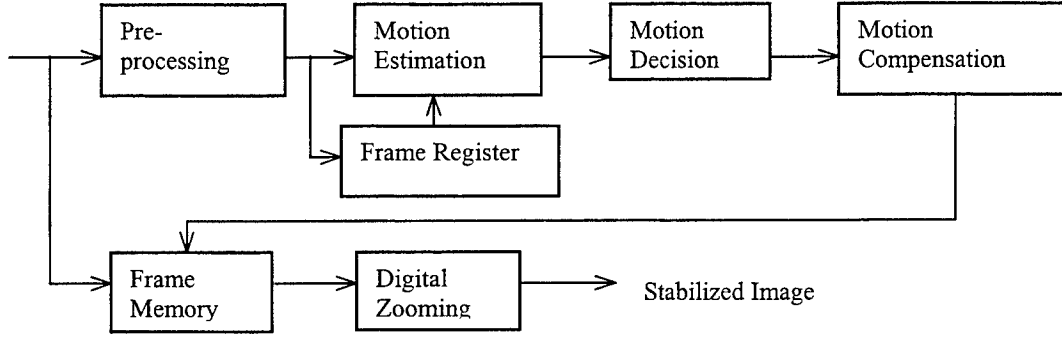
## 1.INTRODUCTION

In recent years, more and more video cameras are accompanied with compact size and powerful zooming capability. The advancements of these features make the image stability problem even more crucial, because an unconscious movement of the holding hand may cause an annoying shaking of the images. Consequently, we usually need a digital image stabilization (DIS) system to soothe the problem. A digital image stabilization system using only image processing techniques could be a suitable solution because such a system can be fully realized in VLSI to fit the compactness requirement. Up to now, many approaches regarding digital image stabilization have been proposed and some of them already have been implemented in video cameras.

Figure 1 shows a typical structure of a digital video camera with a digital image stabilization (DIS) system and a corresponding frame memory (FM) [1]. The frame memory is needed to store current image data and to output the stabilized image data. As shown in Figure 2, a general DIS system usually includes five major components: (1) the pre-processing unit, (2) the motion estimation unit, (3) the motion decision unit, (4) the motion compensation unit for FM, and (5) the digital zooming unit [3].



**Figure 1.** Block diagram of a digital video camera with DIS system.



**Figure 2.** A general structure of DIS system with frame memory

A traditional way to do motion estimation is the block matching method [1][2][3][4][5][6][7]. To reduce computational complexity, these block matching methods usually divide an image into a small number of blocks and select some representative points to calculate the motion vector for each block. Then, they use these block motion vectors to estimate the global motion vector to compensate the movement of the whole image. However, the rough division of an image may cause the loss of local information and the reduction of precision in global motion decision. Without decreasing the accuracy of motion estimation, Sung-Jea Ko and Sung-Hee Lee adopted bit-plane or gray-coded bit-plane block matching to greatly reduce the computational complexity. However, their algorithms are still based on traditional methods for block division, motion decision and motion compensation [8][9]. For these conventional methods, only simple strategies can be applied in motion decision and motion compensation.

To reserve low computational complexity and the high performance of motion estimation, we also use block-matching method over gray-coded bit-planes to do motion estimation. However, we divide a frame into several blocks to do localized block matching for improving the detection of moving objects. We design a new approach, which uses temporal correlation to efficiently detect moving objects and panning conditions. In our architecture, both rotational and translational movements can be compensated. Here, the affine transformation is adopted to align the image contents in different frames. Finally, an efficient and “real-time” hardware is also implemented.

## 2. LOCALIZED BLOCK MATCHING OVER GRAY CODED BIT-PLANE

Sung-Jea Ko and etc. [9] proposed the usage of bit-plane images instead of the 8-bit gray level images. With bit-plane images, the block matching process can be implemented using only binary Boolean functions and thus the computational complexity is significantly reduced. In this paper, we also use gray-coded bit-plane images as the basis to do motion estimation. Assume  $f(x, y)$  is an image and  $g_i(x, y)$  are the corresponding gray coded bit-planes.

That is, if

$$f(x, y) = a_{K-1}(x, y)2^{K-1} + a_{K-2}(x, y)2^{K-2} + \dots + a_1(x, y)2 + a_0(x, y) \quad \text{Eq. (1)}$$

then

$$g_i(x, y) = a_i(x, y) \oplus a_{i+1}(x, y) \quad 0 \leq i \leq K-2 \quad \text{Eq. (2)}$$

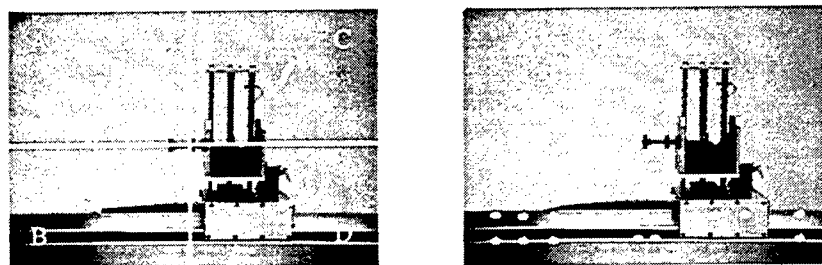
$$g_{K-1}(x, y) = a_{K-1}(x, y)$$

The correlation measure to calculate the motion vectors is defined as:

$$c(m, n) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g'_k(x, y) \oplus g_k^{t-1}(x+m, y+n) \quad \text{Eq. (3)}$$

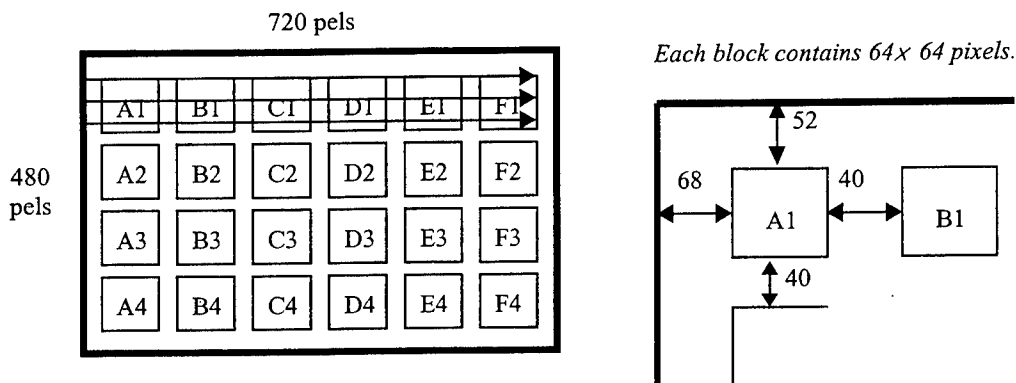
In Figure 3, we demonstrate the comparison of a traditional rough-division method which is working on 8-bit images, and our fine division method which is working on gray-coded bit-planes. Since the operations over gray-coded bit-planes

and much simpler than the operations over 8-bit gray-level images, the computation complexity of our approach is roughly the same as previous approaches even though we have applied a much finer division over the images. The image in Figure 3 is extracted from an image sequence, which is captured by an intentionally shaken video camera. The scene in the image sequence also contains a moving object, which is moving to the right. The traditional rough-division method divides this frame into four blocks and detects four local motion vectors separately: (4,0), (-1,-6), (-5,1), and (-2,2). Three of these 4 motion vectors, except the lower-right block, are not reliable due to either lack of features or the appearance of repeated pattern. The detected motion vector of the lower-right block is also unreliable since this block contains both the motion of a moving object and the motion caused by the shaking camera. Therefore, in this example, the traditional rough-division approach fails to estimate the motion caused by the vibrating camera. On the other hand, with our fine-division approach, many localized motion vectors may still survive and the result is shown in Figure 3(b).



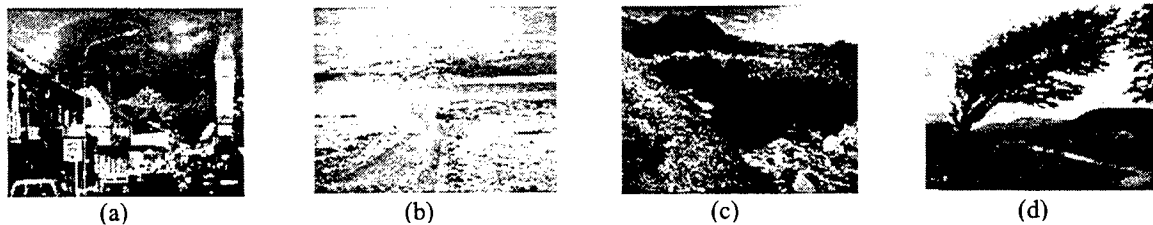
**Figure 3** (a) Rough-division of image and the detected motion vectors.  
(b) Fine-division of image and the detected motion vectors.

As mentioned before, the complexity of the fine-division motion estimation is about the same as before with the computation of a large block being decomposed into a few computations of smaller blocks. The localization of motion estimation has two advantages. First, the presence of some moving objects in the image frame will have less impact on the accuracy of motion estimation. Second, the increased amount of motion vectors may increase the signal-to-noise ratio. However, how to choose a proper block size becomes an important issue. If block size is too small, the accuracy of motion estimation is decreased; while if block size is too large, some local information will get lost. Consequently, for a practical camera system, we choose the block size to be  $64 \times 64$  and we divide each frame into 24 regions, as shown in Figure 4, to meet this trade-off.



**Figure 4.** Illustration of the division of the gray-coded bit-planes in our camcorder system.

To evaluate our approach, we compared the estimated frame motion vectors based on the fine-division over gray-coded bit-plane matching and the conventional method based on the rough-division over 8-bit plane matching. Here we utilize the Root-Mean-Square-Error (RMSE) measure to evaluate the performance.  $RMSE = 1$  means that the average estimation error equals to 1 pixel. We consider four test image sequences ((a)~(d)) as shown in Figure 5 with a resolution of  $640 \times 480$  pixels and 11 frames, which contains simulated hand-shaking motion. The results are display in Table 1. From Table 1(1) and (2), we can observe that the RMSE of our approach is approximate to the conventional method and both of them are far smaller than 1. By observing Table 1(3), while the sequences are suffering in AWGN with variance 0.003, the RMSE is still smaller than one pixel.



**Figure 5.** Four test sequences used for evaluate the performance of motion estimation.

Test image Sequence	(1) 8-bit plane (RMSE)	(2) bit-plane (RMSE)	(3) bit-plane in noise (RMSE)
(a)	0.05391	0.06219	0.11879
(b)	0.06883	0.12458	0.20827
(c)	0.06544	0.12647	0.24014
(d)	0.04002	0.06884	0.07864

**Table 1.** The comparison of RMSE with three different conditions in four test sequences.

### 3.MULTI-RESOLUTION BLOCK MATCHING

To estimate the frame motion vector of  $k$ th frame in an image sequence, we define the  $k$ th frame as the current frame and the  $(k-1)$ th frame as the reference frame. For each block in the current frame, we search within a predefined “searching range” over the reference frame to find the “best” match and thus estimate the localized motion vector of this block. To deal with the possibility of larger movement of hand shaking, a larger searching range is usually needed; however, to reduce the computation complexity, a smaller searching range is preferred. In order to handle the large movement without adding too many computation loads, we adopt a multi-resolution strategy.

With a down-sampling-by-2 multi-resolution structure, the magnitude of the motion vectors in a lower resolution is proximately twice the magnitude of the motion vectors over the corresponding region in the next higher resolution. This phenomenon implies that we may apply the same block matching method over the low-resolution image to estimate the motion vectors when the frame motion is out of the searching range in the higher-resolution image. This multi-resolution approach helps in dealing with large movement of hand shaking.

Besides the detection of large movement, multi-resolution also provides some other advantages. The information of motion vectors estimated in lower resolution could be passed to higher resolution frames. Thus, the searching area in the higher resolution can be well localized into a small area. This will lower the computation load. Furthermore, the motion decision in lower resolution could also be utilized in high resolution. Hence, if we deduce in the lower-resolution layer that a region is lack of texture and its motion vector is unreliable, we may deduce the same conclusion for corresponding region in the higher-resolution layer.

## 4. DETECTION OF EXISTING MOVING OBJECT AND PANNING CONDITION

There exist many factors that may affect the accuracy and performance of motion estimation, which we call "irregular conditions". Many methodologies for detecting these conditions have already been proposed. However, these methods may be very complicated or may not be suitable for our architecture. Consequently, we design our own methodology to detect these conditions for localized block matching over gray coded it-planes. Here, we only concentrate on how we detect moving object and panning conditions.

### 4.1. Random Like Motion and Temporally Correlated Motion

If an image sequence contains a moving object, the regions including this moving object may offer incorrect local motion vector. Thus we need to eliminate these invalid local motion vectors to ensure the accuracy of motion compensation. Here, we propose a method that is efficient and can be easily implemented for detecting the existence of moving object. First, we want to discuss the difference between two kinds of motion: random-like motion and temporally correlated motion. As shown in Figure 6(a), a motion regarded as random-like will fluctuate around zero and the variance of this motion will be relatively large. However, Figure 6(b) shows a temporally correlated motion, which usually moves in a specific direction and the variance of this motion will be relatively small.

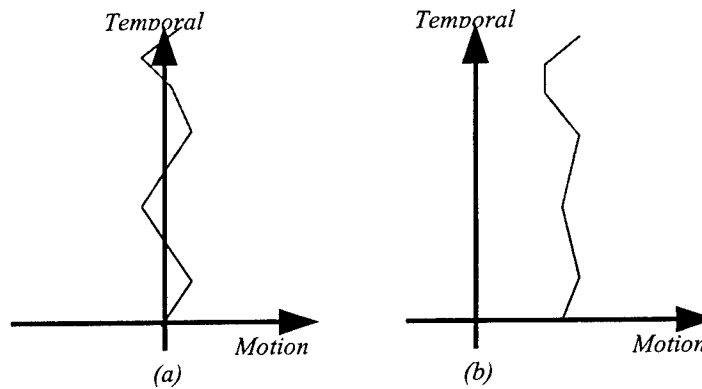


Figure 6. Two kinds of motion: (a) random-like motion (b) temporally correlated motion.

### 4.2. Existing Moving Object and Panning Condition

After we have discussed the major difference between random-like motion and temporally correlated motion, we may find that these two types of motion are closely related to the motion caused by hand shaking and the motion caused by intentional panning. The motion caused by hand shaking makes the captured scene fluctuate around the center of focus. This makes the motion vectors fluctuate around zero. On the other hand, for the intentional motion, like panning, tends to move in the same direction for a short time. Consequently we classify the motion caused by hand shaking as random-like motion and the motion caused by intentional panning or the existence of moving object as temporal correlated motion. Here we design a simple test, as shown below, to distinguish these two kinds of motion:



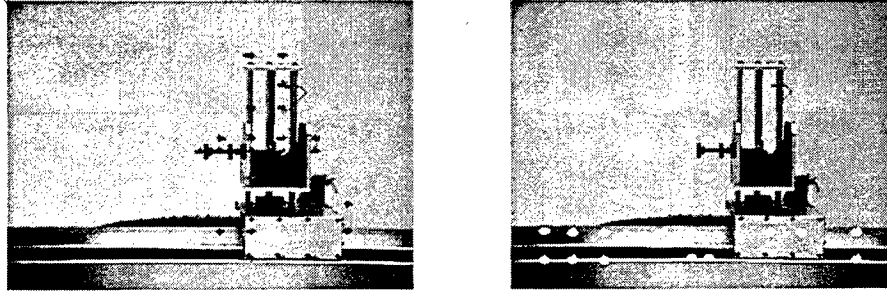
$$|MV(t_1) - MV(t_2)| + |MV(t_2) - MV(t_3)| + \dots + |MV(t_{N-1}) - MV(t_N)| = T_1 \quad \text{Eq. (4)}$$

$$\frac{1}{N} \sum_{i=1}^N MV(t_i) = T_2 \quad \text{Eq. (5)}$$

*If  $T_1/T_2 < K_1$  and  $T_2 > K_2$  then temporal correlated motion  
else random-like motion*

In this test, we observe the frame motion vector along the temporal domain. Assume  $MV(t_1)$  denotes the frame motion vector at time  $t_1$  and  $MV(t_N)$  denotes the frame motion vector at time  $t_N$ , the end of observation. In our simulation, we choose  $N=8$ ,  $K_1=5$ , and  $K_2=1$ . If a motion behaves as temporally correlated motion, its variance (similar to  $T_1$ ) is usually small and its mean ( $T_2$ ) is usually large. Figure 7 shows the experiment result. The test sequence contains two motions: (a) a temporally correlated motion at the slider and (b) a random-like motion for the remaining part. The simulation shows that these local motion vectors detected as temporally correlated motion are locating around the slider.

+



**Figure 7.** The simulation results: (a) temporal correlated local motion vectors (b) random-like local motion vectors.

After the temporally correlated motion vectors are localized, we use them as the clues of existing moving objects. However, if the temporal correlated motion vectors are globally present, the camera is under an intentional panning. Figure 8 shows the test sequence. This test sequence including a walking lady was captured under an intentional panning movement. After temporally correlated test and globally correlated test, we detect that this sequence is under an intentional panning.



**Figure 8.** The test sequence with an intentional panning.

## 5. MOTION COMPENSATION WITH AFFINE MOTION MODEL

Affine transform is a popular way to describe linear motion, rotation, and some deformation. Motion composed of not only translation but also rotation is a very common and can be modeled by using affine motion model. Equation (6) shows the equations of affine transform.

$$\begin{cases} \overline{X}_{t+1} = aX_t + bY_t + c \\ \overline{Y}_{t+1} = dX_t + eY_t + f \end{cases} \quad \text{Eq.(6)}$$

$(\overline{X}, \overline{Y})$ : the coordinates of the compared frame  $(X, Y)$ : the coordinates of the reference frame

To estimate the six parameters (a~f) in the affine model, we use the least mean square method. Assume there are N valid motion vectors. We use the standard optimization method to find the “optimal” coefficients that may minimize the following equations:

$$\begin{aligned} & \sum_{n=1}^N (aX_n + bY_n + c - \overline{X}_n)^2 \\ & \sum_{n=1}^N (dX_n + eY_n + f - \overline{Y}_n)^2 \end{aligned} \quad \text{Eq.(7)}$$

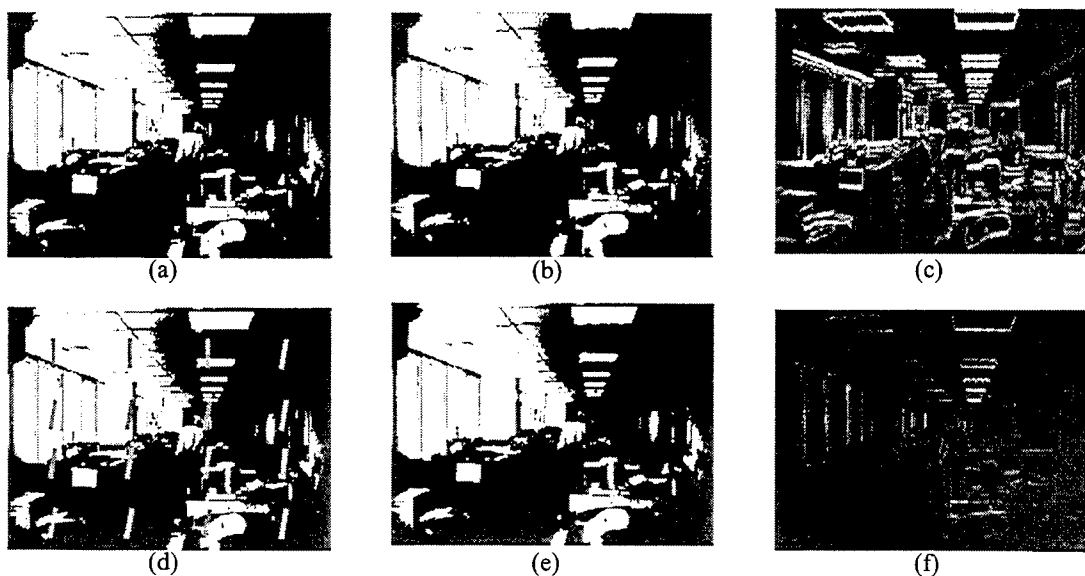
Equation (8) shows the detail how we calculate the coefficients a~f :

$$\begin{aligned} \begin{bmatrix} a \\ b \\ c \end{bmatrix} &= \begin{bmatrix} X_1^2 + X_2^2 + \dots + X_n^2 & X_1Y_1 + X_2Y_2 + \dots + X_nY_n & X_1 + X_2 + \dots + X_n \\ X_1Y_1 + X_2Y_2 + \dots + X_nY_n & Y_1^2 + Y_2^2 + \dots + Y_n^2 & Y_1 + Y_2 + \dots + Y_n \\ X_1 + X_2 + \dots + X_n & Y_1 + Y_2 + \dots + Y_n & n \end{bmatrix}^{-1} \times \\ & \begin{bmatrix} \overline{X}_1X_1 + \overline{X}_2X_2 + \dots + \overline{X}_nX_n \\ \overline{X}_1Y_1 + \overline{X}_2Y_2 + \dots + \overline{X}_nY_n \\ \overline{X}_1 + \overline{X}_2 + \dots + \overline{X}_n \end{bmatrix} \\ \begin{bmatrix} d \\ e \\ f \end{bmatrix} &= \begin{bmatrix} X_1^2 + X_2^2 + \dots + X_n^2 & X_1Y_1 + X_2Y_2 + \dots + X_nY_n & X_1 + X_2 + \dots + X_n \\ X_1Y_1 + X_2Y_2 + \dots + X_nY_n & Y_1^2 + Y_2^2 + \dots + Y_n^2 & Y_1 + Y_2 + \dots + Y_n \\ X_1 + X_2 + \dots + X_n & Y_1 + Y_2 + \dots + Y_n & n \end{bmatrix}^{-1} \times \\ & \begin{bmatrix} \overline{Y}_1X_1 + \overline{Y}_2X_2 + \dots + \overline{Y}_nX_n \\ \overline{Y}_1Y_1 + \overline{Y}_2Y_2 + \dots + \overline{Y}_nY_n \\ \overline{Y}_1 + \overline{Y}_2 + \dots + \overline{Y}_n \end{bmatrix} \end{aligned} \quad \text{Eq.(8)}$$

It seems that Eq.(8) is a little complicated for a practical implementation. Nevertheless, note that all the elements in the matrices can be treated as the inner product of two vectors and some of these entries are duplicated. This implies that this computation can be efficiently implemented with a fast algorithm of vector inner product. Figure 9 shows the simulation of motion compensation after using the affine model. Figure 9(a), (b) illustrate two consecutive image frames with a rotation motion. Figure 9(d) shows the valid local motion vectors after motion estimation and motion compensation. Based on these valid motion vectors, we calculate the coefficients of affine transform and Figure 9(e) shows the stabilized image frame. Figure 9(c) and 9(f) shows the intensity difference before and after motion compensation.

## 6.CONCLUSION

In this paper, we design a fine division method for block matching over gray-coded bit-planes to acquire high performance of stabilization. We also design our new strategies to efficiently detect moving object and intentional panning by using the test of random like motion and temporally correlated motion. The affine transform is used for motion compensation to model camera motion with rotation. Based on this architecture, a “real-time” motion estimation hardware interconnecting with a microprocessor is designed too, and it has already been implemented on FPGA 10K 100.



**Figure 9.** The simulation results: (a) the reference frame (b) the compared frame (c) difference between (a) and (b) (d) valid local motion vectors (e) the aligned frame without interpolation (f) difference between (a) and (e).

## REFERENCE

- [1] Joon Ki Paik, Yong Chul Park, and Dong Wook Kim, "An Adaptive Motion Decision System for Digital Image Stabilizer Based on Edge Pattern Matching", *IEEE Trans. on Consumer Electronics*, Vol. 38, No.3, AUGUST 1992.
- [2] Kenya Uomori, Atsushi Morimura, Hirofumi Ishii, Takashi Sakaguchi, and Yoshinori Kitamura, "Automatic Image Stabilizing System by Full-digital Signal Processing", *IEEE Trans. on Consumer Electronics*, Vol. 36, No.3, AUGUST 1990.
- [3] Toshiro Kinugasa, Naoki Yamamoto, and Hiroyuki Komatsu, "Electronic Image Stabilizer for Video Camera Use", *IEEE Trans. on Consumer Electronics*, Vol. 36, No. 3, AUGUST 1990.
- [4] Yo Egusa, Hiroshi Akahori, Atsushi Morimura, and Noboru Wakami, "An Application of Fuzzy Set Theory for an Electronic Video Camera Image Stabilizer", *IEEE Trans. on Fuzzy Systems*, Vol. 3, No.3, AUGUST 1995.
- [5] Yo Egusa, Hiroshi Akahori, Atsushi Morimura, and Noboru Wakami, "An Electronic Video Camera Image Stabilizer Operated on Fuzzy Theory", IEEE 1992.

- [6] Carlos Morimoto, and Rama Chellappa, "Evaluation of Image Stabilization Algorithms", IEEE 1998.
- [7] Masayoshi Sekine, Toshiaki Kondou, and Hisataka Hirose, "Motion Vector Detecting System for Video Images Stabilizers", *IEEE Trans. on Consumer Electronics*, 1994.
- [8] Sung-Hee Lee, Kyung-Hoon Lee, and Sung-Jea Ko, "Digital Image Stabilizing Algorithms Based on Bit-plane Matching", IEEE 1998.
- [9] Sung-Hee Lee, Seung-Won Jeon, Eui-Sung Kang and Sung-Jea Ko, "Fast Digital Stabilizer based on Gray Coded Bit-Plane Matching", *IEEE Trans. on Consumer Electronics*, Vol. 45, No.3, AUGUST 1999.
- [10] Jung-Hyun Hwang, Hweihn Chung, Sung-II Su, Yong-Chul Park, and Chul-Ho Lee, "High Resolution Digital Zoom Using Temporal IIR Filter", *IEEE Trans. on Consumer Electronics*, Vol. 424, No.3, AUGUST 1996.
- [11] Joon Ki Paik, Yong Chul Park, and Sung Wook Park, "An Edge Detection Approach to Digital Image Stabilization Based on Tri-state Adaptive Linear Neurons", *IEEE Trans. on Consumer Electronics*, Vol. 37, No. 3, AUGUST 1991.
- [12] M. Hansen, P. Anandan, K. Dana, G. van der Wal, and P. Burt, "Real-time Scene Stabilization and Mosaic Construction" in *The 1994 Image Understanding Workshop*, Nov. 1994.

## **SESSION 4**

### **Color Imaging**

# Internet Color Imaging

Hsien-Che Lee

Imaging Science and Technology Laboratory  
Eastman Kodak Company, Rochester, New York 14650-1816 USA

## ABSTRACT

The sharing and exchange of color images over the Internet pose very challenging problems to color science and technology. Emerging color standards will solve many of the problems we face today, but existing images of unknown origin and output devices of unknown calibration will continue to cause problems for many users. This paper presents a brief overview of available solutions to some of the problems and suggests some directions for future research. Although most existing solutions are quite primitive and fragile, the rapid advance of computing technology promises to bring more sophisticated and intelligent image processing algorithms to common practical use. Image understanding, scene physics, visual calibration, and image perception are four areas of research that are beginning to make good progress toward a fully automatic quality optimization system for color imaging applications.

**Keywords:** Internet, color imaging, automatic calibration, scene statistics, digital image processing

## 1. INTRODUCTION

When we order a sweater from a web site, how do we know if the color of the sweater is what we like? When we send a digital camera image to an on-line fulfillment center, how do we know if they can print it with good tone and color rendition? When we download a color image from a web site and print it on the color printer at our home, how can we make it come out as beautiful as we see it on our color monitor? These are questions that we are facing everyday. Color imaging applications for Internet shopping, services, and information gathering have become ubiquitous. Yet, color images that are exchanged over the Internet originate from widely varied sources, display/printing devices used to show those images are not calibrated, and viewing conditions are rarely controlled. So how can the whole thing work? Well, chaotic as it may be, there are several factors that save us from a total breakdown in such a mess: (1) Our visual system is very capable and very forgiving. With an amazing grace, it can often adjust to the distortion and extract the information needed. It is not that we do not see the distortion, it is that we choose not to pay too much attention to it. (2) Devices are built to vaguely conform to various standards, and those standards are not drastically different from each other. (3) We don't know what we are missing. Occasionally we see great pictures on our monitors and we are pleasantly surprised. We rarely ask, why don't we get great pictures all the time?

There are three basic classes of technical problems in Internet color imaging: (1) color images of unknown calibration, (2) imaging devices of unknown characteristics, and (3) viewing conditions of unknown perceptual effects. Solutions to each of these problems vary from completely manual to fully automatic adjustments, from closed systems to standardized interfaces, and from approximation to perfection. These problems and their possible solutions are discussed in this paper, and future research directions are suggested in the discussion.

## 2. STANDARDIZATION

The major component in the solution of most problems in Internet color imaging is to standardize the protocols of how color information should be communicated. The protocols have to be complete in all technical details. For example, it is not sufficient to specify the RGB signals from a digital camera as gamma-corrected video signals. Ideally, the spectral response functions of the camera should be provided with the camera. However, most consumers do not know how to take advantages of this type of technical information, or are unwilling to spend the time to do so. Therefore, national and international standardization efforts are mostly directed toward simple and packaged solutions. So instead of asking for the manufacturers to provide the technical information with the products, standards tend to describe a recommended system and its signal specifications, and it is up to

each manufacturer to produce products that can work “reasonably well” with the standard system. This is a very practical and inexpensive solution, although it means that the needed technical information is often not available to the knowledgeable consumers. For example, chromaticity coordinates of the phosphors of a CRT monitor are usually not provided to the user.

Among the various international standard bodies, ISO, IEC, CIE, and ITU are four of the major organizations that publish standards of direct interest to the field of color imaging. The International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), and the International Commission on Illumination (CIE) are the three major organizations that establish voluntary international standards. The International Telecommunication Union (ITU) is organized by the United Nations and its standards are regulatory through government administrations and treaties.<sup>1,2</sup> These international standards are then, in turn, used by industries to set up proposals for other applications. For example, the ITU-R Recommendation BT.709 (Parameter Values for the HDTV Standards for Production and International Programme Exchange) and Recommendation BT.1200 (Target Standards for Digital Video Systems for the Studio and International Programme Exchange) are two international standards that are widely adopted and adapted in color imaging applications, such as KODAK PHOTOYCC Color Interchange Space<sup>3</sup> and sRGB<sup>4</sup> color encodings.

In order to facilitate the standardization of color management systems, a non-profit organization, called International Color Consortium (ICC), was established in 1993.<sup>5</sup> The basic idea is to provide a device profile for every imaging device so that color data produced by one device can be translated into a device-independent profile connection space (PCS), which can, in turn, be converted into the native color space of another device. The ICC profile format is described by the document published by the Consortium. Although the interpretation of the rendering intent of some profile parameters can be ambiguous,<sup>6</sup> the ICC profiles, if fully implemented by most imaging hardware and software companies, will be a giant step forward toward solving many (but not all) problems in Internet color imaging applications. However, for adjustable devices, such as monitors, scanners, and digital cameras, a fixed device profile is obviously not sufficient. For example, if the contrast or brightness knob of a monitor is adjusted by the user, the monitor device profile is no longer valid for the status of that monitor.

A major advantage of the device profile approach to color management is that color images can be saved in the native color space of the device without unnecessary quantization to an intermediate color space.<sup>7</sup> This is especially important for 8 bits/color/pixel images. A fundamental problem with color solutions based on standards is that the color images are at best colorimetrically or perceptually correct (remember, this is a great position to be in), but may be far from visually optimum in quality. This is partially due to the fact that standards are related to systems and devices, not individual images. It is also partially due to our lack of understanding in human perception.

### 3. IMAGES OF UNKNOWN CALIBRATION

Most color images existing on Internet do not have any calibration information associated with them. Similarly, many images sent to on-line fulfillment centers are not calibrated. How do we deal with such images?

#### 3.1. Interactive Mode

If a human operator is engaged in processing such images, a good strategy is to first estimate its basic tone scale metric. Are the digital numbers proportional to linear luminance, log luminance, or video (gamma corrected) luminance in the scene? These three are the most often used metrics from CCD sensors, film, and video cameras. We can make the assumed transform (with some variations in parameters) and display the image on a calibrated monitor to see which of these transforms look best and take a bet. However, most color images are produced through some nonlinear tone scale curves. Therefore, there will be a lot of work to adjust the highlight and the shadow to make the image look right with our intended tone reproduction curve. Because many Internet images are meant to be viewed on CRT monitors, they are very likely to be in gamma-corrected video space (such as NTSC-RGB or sRGB). The next step is to extract the digital color values from what we think are the neutral (gray or white) objects and the skin areas. The neutral objects will help us to do a basic color balance. The skin areas will allow us to estimate the color matrix required to rotate the color axes to the ones we want to use in our device. This is easier to say than done. Although the unexposed skin area of a given race tend to have a certain reflectance value, the exposed skin areas tend to vary greatly in lightness. Table 1 shows some sample

**Table 1.** Sample measurements of skin color (forehead and cheek).

African	$L^* = 37.6 \pm 1.3$	$a^* = 6.9 \pm 1.4$	$b^* = 10.7 \pm 2.3$
Arabian	$L^* = 61.5 \pm 2.3$	$a^* = 5.6 \pm 1.1$	$b^* = 17.3 \pm 1.8$
Caucasian	$L^* = 66.3 \pm 2.8$	$a^* = 11.2 \pm 0.9$	$b^* = 12.3 \pm 1.8$
Japanese	$L^* = 60.7 \pm 4.37$	$a^* = 10.8 \pm 2.36$	$b^* = 17.1 \pm 2.19$
Vietnamese	$L^* = 65.1 \pm 3.1$	$a^* = 5.4 \pm 0.8$	$b^* = 15.4 \pm 1.1$

measurements of (forehead/cheek) skin colors.<sup>8-10</sup> From the spectral measurement data reported by Edwards and Duntley,<sup>11,12</sup> we have computed the tristimulus values of the skin color of various races. In general, the dominant wavelength of (unexposed) skin is relatively constant across races (at about 584 nm under D<sub>65</sub> illuminant). The main difference is in the luminance factor (varies from 7% to 45%) and the excitation purity (varies from 17% to 33% under D<sub>65</sub>). The effect of sun tan is to shift the dominant wavelength toward a longer wavelength by an amount on the order of 10 nm, while the excitation purity remains about the same. Knowledge of typical distributions of skin color only gives us some estimates of how much and which way a color correction should be given. In the interactive mode, an operator can look at the image on a calibrated monitor and make continuous adjustments as needed.

### 3.2. Automatic Mode

For many applications, the cost and throughput requirements cannot afford too much operator intervention. In these cases, automatic algorithms have to be developed so that tone, color, and sharpness correction can be carried out all by computers, on an image by image basis. Automatic correction requires that the image calibration be estimated first and then the desired manipulation applied. The first step is the more difficult one of the two and currently, there is no good solution. However, there are potential research directions that we can see from some of the existing approaches.

In general, it is not possible to derive the exact relationships between the scene radiances and the digital values in a given digital image. However, it is interesting to note that when we display or print an image with a wrong calibration, we can often see from the rendered image that something is not quite “right”. How can we sense that? What is it in the rendered image that is telling us that something is not right? Are there some “invariant” or “inherent” features in natural scenes that our visual system has learned to recognize and when those features are not reproduced well in an image because of wrong calibration, our visual system will sense the distortion of those features? It is difficult to imagine that such “invariant” features can exist. However, several studies have shown that indeed some characteristic features do exist for natural scenes. For example, the amplitude  $A$  of the radial spatial frequency  $f$  of natural scenes tend to be a power function<sup>13-15</sup> of the frequency  $f$ :  $A(f) = \alpha f^{-p}$  and typical value of  $p$  is between 0.8 and 1.2. Because this characteristics is found to be relatively insensitive to calibration,<sup>15</sup> it is not useful for estimating the calibration from an image.

#### 3.2.1. Tone correction

One of the features that has been proposed<sup>16</sup> for estimating the unknown tone scale calibration is the statistical property that the log-exposure distribution of a natural scene is approximately a Gaussian distribution. It is argued<sup>16,17</sup> that this property is due to the random distributions of surface orientation, reflectance factors, textures, and illumination, and also partially due to the central limit theorem. One interesting observation from the underlying heuristic reasoning is that the theoretical distribution holds true for any spectral response function used to measure exposure. This can be used for color correction for images that have mixed illuminants.<sup>16</sup> However, it is very easy to give counter examples in which such a Gaussian property does not hold for individual images or even for an ensemble of images,<sup>18</sup> depending on the contents on the images. For example, if we take an outdoor picture that includes a significant portion of the sky, the log-exposure distribution of the image is most likely to be bimodal. One might still argue that each mode of the histogram can be approximated by a Gaussian distribution. Unfortunately, even a mixture of Gaussians is often not a good model, because if there are one or more large uniform areas in the image, the shape of the log-exposure distribution will be quite varied. One way

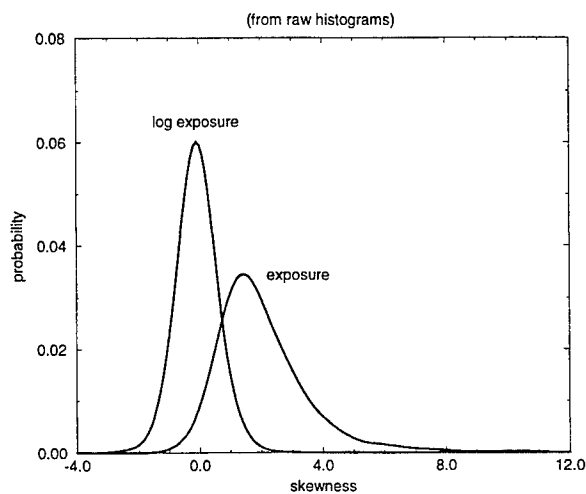


to reduce the bias introduced by a large uniform area is to sample only where color or exposure signals change significantly.<sup>16</sup> The other way is to allow deviation from normality in a parameterized family of distributions.<sup>19</sup>

The problem of estimating the unknown calibration can be greatly simplified if we are interested in classifying the unknown input into one of the three most widely used metrics: linear-exposure, log-exposure, and video gamma-corrected exposure. For example, a simple way to classify images of unknown calibration is to take advantage of the fact that the histogram of a log-exposure image is more symmetric with respect to its mean than a linear-exposure image. If the histogram of the image in question is highly skewed to the right, it is more likely to be a linear-exposure image. The skewness of a distribution  $h(x)$  can be measured by:

$$\text{skewness} = \frac{m_3}{\sigma^3}$$

where  $m_3$  is the third central moment, i.e.,  $m_3 = \int (x - \mu)^3 \cdot h(x) dx$ , and  $\mu$  and  $\sigma$  are the mean and the standard deviation. We have calculated the skewness of the linear-exposure histograms and that of the log-exposure histograms for 1800 consumer images. Figure 1 shows the skewness distributions for these two image metrics. The two distributions intersect at skewness = 0.75. The fraction of log-exposure images that have a skewness

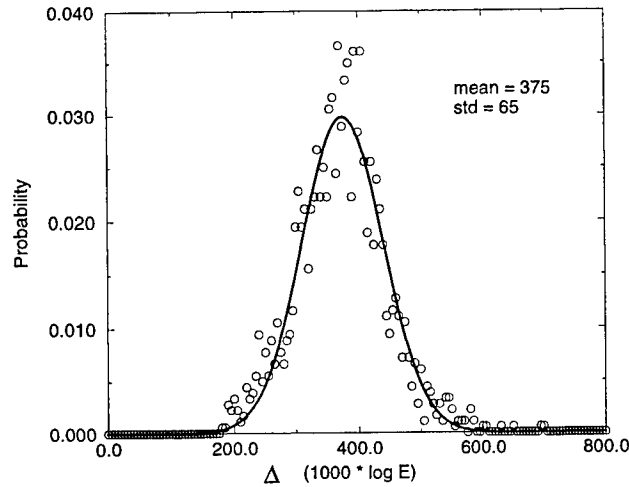


**Figure 1.** The distributions of skewness for exposure and log-exposure histograms for 1800 consumer images. (The two distribution curves have been smoothed.)

greater than 0.75 is about 11.2%. The fraction of linear-exposure images that have a skewness less than 0.75 is about 16.3%. Thus, from the skewness of the histogram alone, we can classify the input image into one of the two metrics (linear-exposure or log-exposure) correctly more than 80% of the time. In fact, we can improve the classification by using the skewness of the histogram accumulated only from the edge pixels, thus excluding the biases from large uniform areas. If we only sample along edge pixels in the images and calculate the skewness of the log exposure histograms of the edge pixels, we find that the standard deviation of the skewness distribution is now reduced from 0.59 to 0.42. The mis-classification rate of rejecting a log-exposure image has dropped to 3.8%. However, the above experimental results are based on the two-class discrimination problem. The algorithm does not work well when we have to deal with the three-class problem for linear, log, and video metrics.

Suppose that we have successfully estimated the unknown metrics of the input image, what can we do to improve the tone rendition of the image? This is a much easier problem. Although really robust algorithms are still being developed, several methods exist for adjusting global or local contrast of an image. For example, a global contrast adjustment algorithm<sup>19</sup> can take advantage of pre-compiled statistical data for some scene contrast estimator, such as the standard deviation,  $\Delta$ , of the histogram of edge pixels. We have compiled such statistics. Figure 2 shows the distribution of  $\Delta$  calculated for 1800 consumer images. Its mean is 0.375 in log-exposure. If we take  $\pm 3 \Delta$  (i.e. 6 standard deviations of the log-exposure histogram) as the dynamic luminance range of

the scene, we have an average log scene luminance range of  $0.375 \times 6 = 2.25$ , which corresponds to a luminance range of 168:1. This number is indeed very close to the average scene luminance range of 160:1 reported by Jones and Condit in their classical study using actual measurements on many natural scenes.<sup>20</sup> Another study<sup>21</sup> also reported that the average standard deviation of the log-exposure histograms is 0.33. The mutual confirmation of these studies does not mean that the current contrast estimate is accurate, but it shows that the algorithm can produce a reasonable result with a very simple computational procedure that does not require much prior knowledge. Further experimental testing is needed to achieve the optimal contrast adjustment. From Fig. 2 one



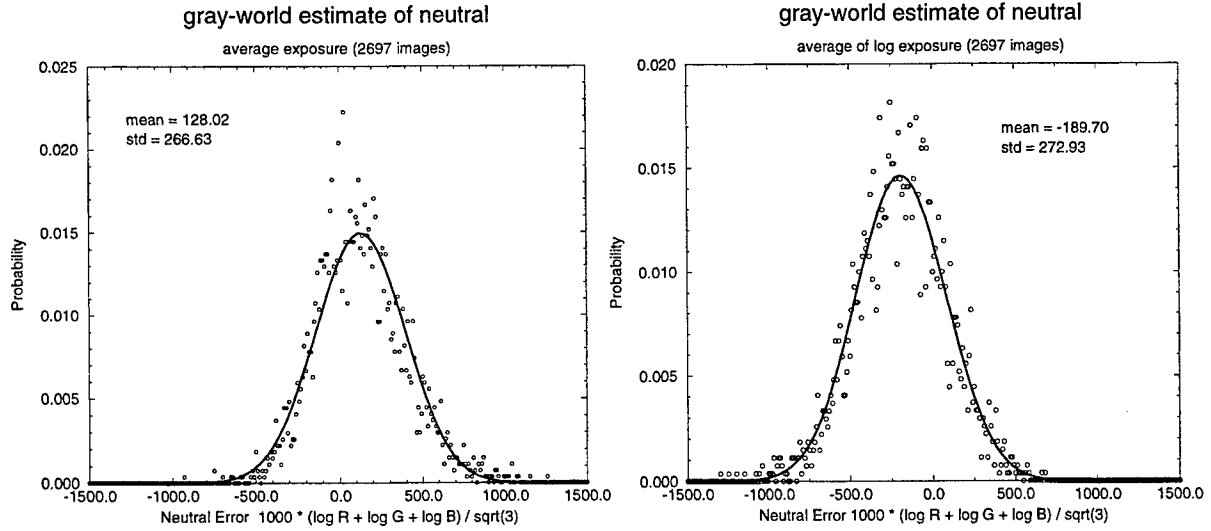
**Figure 2.** The distribution of the standard deviation,  $\Delta$ , of the log-exposure histogram of the edge pixels of an image for 1800 consumer images.

can see that a scene dynamic range can be as high as  $0.55 \times 6 = 3.3$  in log-exposure (or about 2000:1 in exposure) and as low as  $0.2 \times 6 = 1.2$  in log-exposure (or about 16:1 in exposure). For most images of small dynamic range, say less than 80:1, experimental results so far show that the contrast adjustment greatly improves their perceived image quality.

In addition to contrast adjustment, it is also necessary to decide how light or how dark an image should be display or printed. This problem is called the density balance problem in photofinishing and it is similar to the exposure control problem in digital camera design. The problem is stated as follows: given a digital image, determine the digital value that is to be displayed at a given luminance level or printed at a given reflectance, so that the rendered image looks optimal in tonal quality. Most existing algorithms for density balance are proprietary and not available in public domain. A well-known algorithm is the integration-to-gray method<sup>22</sup> and its many variations. Using a consumer image database in which the optimum balance point (aim) for every image was determined by three experts, we have tested how well the simple integration-to-gray method works on consumer images. The database consists of 2697 images collected from consumers. The integration is done in two ways: averaging in exposure and averaging in log-exposure. The integrated red, green, blue values are used to compute a neutral balance point by the following equation:

$$L = \frac{1}{\sqrt{3}}(\log R + \log G + \log B). \quad (1)$$

This computed balance point is then compared with the experts' optimum point. Figure 3 shows the error distributions along this neutral "luminance" axis. There are two interesting observations from these statistical data: (a) Averaging in exposure and averaging in log-exposure yield about the same magnitude of estimation error. (b) Averaging in exposure and averaging in log-exposure have the opposite bias - the averaged exposure is lower, while the averaged log-exposure is higher than the experts' aim. In general, the gray world assumption produces a mediocre estimate for density balance. One of the obvious problem of the integration-to-gray method is that if there are large uniform areas in an image, they will bias the average to the luminances of those areas.



**Figure 3.** Comparison of Neutral error distributions. Left: averaging in exposure; Right: averaging in log-exposure.

Again, as we discussed before, an obvious improvement is to sample only on edge pixels<sup>16,23</sup> or from active (busy) image regions.<sup>24,25</sup>

### 3.2.2. Color correction

Similar to tone correction, there are two steps in color correction. The first step is to estimate the unknown color calibration and the second step is to apply the desired color manipulation (such as color balance) and enhancement (such as boosting color chroma).

In tone correction, the estimation of unknown calibration is mainly for deriving the functional relationship between the scene radiance and the digital image value. Usually, there is no explicit attempt to estimate how to combine the red, green, blue image values to produce what would be measured by the CIE luminous efficiency function  $V(\lambda)$ . The reason is that in tone correction, we are mainly interested in the intensity relationship between exposure and image value, rather than spectral relationship. When we deal with color correction, the spectral relations become important. It is no longer sufficient for us to know that our image values are proportional to linear-exposure or log-exposure. We actually have to know how they are related to the colors we see. Let  $R, G, B$  be the red, green, blue exposures of a pixel in an image and let  $X, Y, Z$  be the tristimulus values of the object color corresponding to that pixel. The simplest approximation of the relationship between  $R, G, B$  and  $X, Y, Z$  is a  $3 \times 3$  matrix,  $M$ , i.e.,

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = M \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (2)$$

Theoretically, a  $3 \times 3$  matrix is an exact transformation if the spectral response functions of the image capture system are linear combinations of the CIE color matching functions. Because most imaging systems are far from that, a  $3 \times 3$  matrix may not be a good approximation for our images at all. However, currently, this simple approximation is as complicated as we can try to estimate.

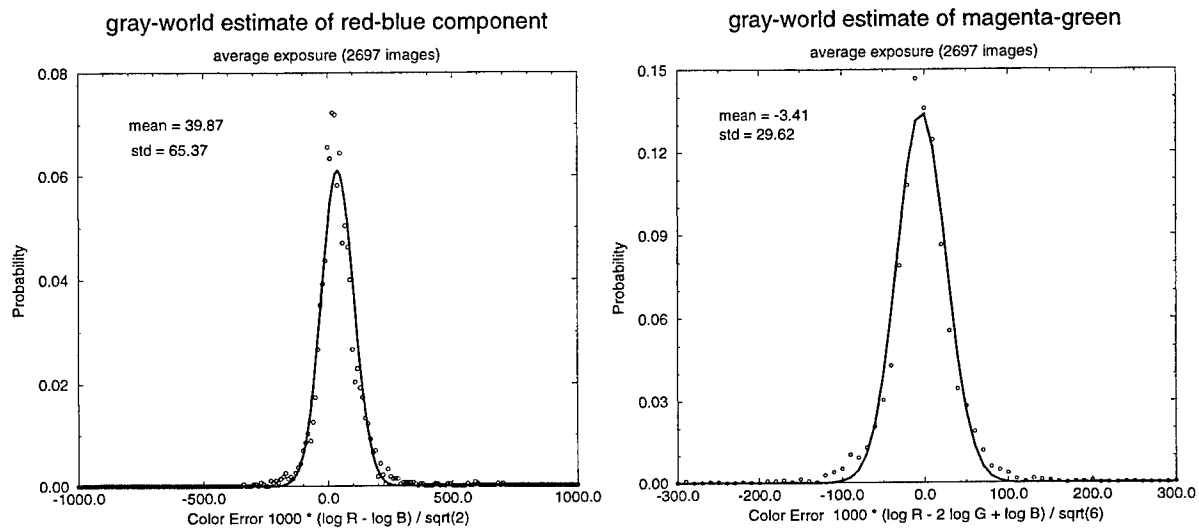
There are nine unknown elements in  $M$  to be estimated. Since in the image capture and printing processes, an overall scale factor can be and will be adjusted on an image by image basis. This is solved as the density balance problem we discussed in tone correction. The remaining eight unknowns can be determined from four pairs of corresponding chromaticity coordinates in  $(R, G, B)$  and  $(X, Y, Z)$ . So, which four possible pairs can we estimate from an image automatically? Two important pairs are the neutral (gray) color and the skin color. The problem of estimating the neutral color in the image is called the color balance problem. The existing algorithms for solving the problem have been reviewed elsewhere.<sup>26</sup> Despite many new algorithms developed for

color constancy, the gray world assumption continues to be the backbone of the color correction algorithms for most printers and video cameras. But, just how gray is the world? If we average the exposure of all the pixels in a color image, we obtain 3 numbers: the average red, green, and blue values, which can be represented as a point in the three-dimensional (R,G,B) color space. In order to remove the exposure differences among images, the R,G,B aims (established by expert judges) of that image are subtracted from the image averages, so that if the averages predict the aims perfectly, the point representing the image should fall at the origin of the (R,G,B) color space. If we do this for 2697 images, we obtain a cluster of points, each representing an image. In order to show the error distribution from the gray world assumption, we project the errors to the red-blue direction and the magenta-green direction, because they are close to the eigenvectors of the covariance matrix computed from all the pixels in the 2697 image. The two chromatic axes are defined as:

$$s = \frac{(\log R - \log B)}{\sqrt{2}} \quad [\text{red} - \text{blue}]$$

$$t = \frac{(\log R - 2 \log G + \log B)}{\sqrt{6}} \quad [\text{magenta} - \text{green}]$$

Figure 4 shows how the errors are distributed. As can be seen from these figures, the error distributions tend



**Figure 4.** Comparison of red-blue and magenta-green error distributions. Left: error distribution in the red-blue direction; Right: error distribution in the magenta-green direction.

to have higher peaks and wider tails than a Gaussian distribution with the same mean and standard deviation. The gray world estimation of color balance point is clearly much better than its corresponding estimate for the density balance point. The standard deviations from the aim values are much smaller, compared with that shown in Fig. 3.

For detecting skin colors, there are two main approaches. One approach<sup>27-29</sup> is to compile the statistical distribution of skin pixels and use it with other shape and texture cues to decide if a pixel or a region in a new input image belongs to skin color. The other approach is to detect human faces in the image.<sup>30-34</sup> However, as we mentioned before, detecting skin regions does not give us a unique chromaticity pair because skin chromaticities are functions of race, sun tan, scene illumination, and many other factors. Regional, seasonal, and cultural statistics can give us some prior distribution of skin chromaticities to help the algorithms make the best estimates.

Given the neutral and skin colors, we still need two more pairs of chromaticities before we can estimate the matrix *M*. Other candidate colors are sky, soil, and grass. Unfortunately, their natural chromaticities are even more varied than the skin color. For outdoor scenes, a possible color vector is the daylight locus. It has been shown<sup>35</sup> that for a color imaging system whose spectral response bands are not too wide (say, on the order of 100

nm), the chromaticity distribution of a color image tend to be elongated along the natural daylight locus. This distribution tendency can also be seen in the data reported in other studies.<sup>13</sup> This is mainly due to the mixed illumination of sunlight and skylight on object surfaces. Because the chromaticity distribution of any given color image is heavily biased by the content of the scene, this daylight characteristic can be used only when many images from the same imaging system are available. In practice, this is not an unreasonable constraint because customer orders tend to come in film rolls or image groups.

### 3.2.3. Image enhancement

Image capture and display/printing processes invariably introduce blur and noise into the images. Image sharpening and noise suppression are two image enhancement operations that have been studied for many years.<sup>36</sup> New algorithms<sup>37-39</sup> using wavelet transforms are also becoming very promising.

In order to sharpen an image and suppress the noise, it is most desirable to have methods for estimating how much and what type of sharpening is needed, and for estimating the noise level as a function of signal in the image. Image blur caused by object motion, focus error, camera optics, film, and scanner can be a complex function to model.<sup>40</sup> In consumer images, image blur is usually not too serious in the sense that most edges are still detectable. An intuitive approach for estimating image blur is to detect all high-contrast, straight edges in the image. By certain heuristic criteria (such as chromatic edges<sup>41</sup> and contrast-normalized gradient<sup>42</sup>), we can locate physical edges that are likely to be straight occlusion edges. The blur function can then be estimated from the edge profiles.<sup>43</sup> Alternatively, edge blur can be modeled and the model parameters estimated from the profiles.<sup>44</sup>

Noise estimation has been studied many times<sup>45,42,44</sup> in the past. A rough estimate of homogeneous, signal-independent white noise is not difficult to compute whenever the image contains some uniform area. However, when the entire image is full of busy textures, all existing methods seem to fail. Fortunately, most consumer images have some uniform area if local shading is removed by polynomial fitting.

## 4. DEVICES OF UNKNOWN CHARACTERISTICS

In order to achieve good tone and color reproduction, all imaging devices should be carefully calibrated. However, color calibration requires expensive instruments, technical knowledge, and time-consuming efforts. Therefore, most monitors and printers used at home and offices are not calibrated at all. As a consequence, images are typically displayed or printed at less than desirable quality. The chaotic situation is mainly caused by the lack of well accepted standards. The other major contributor is the stability of most imaging devices, whose characteristics change with time, temperature, humidity, usage, and other uncontrollable factors. These two major causes of chaos can be dealt with by consensus of default standards and by development of easy to use tools for characterizing imaging devices either with inexpensive instruments or with visual judgment.

### 4.1. Default Standards

Standards are driven by competing forces and that is why they are often compromised solutions. However, no standard is worse than a sub-optimum standard. The other driving force is the speed of technology development. It means that trying to perfect a standard may take longer than the life of the current technology.

The ITU-R Recommendation BT.709 forms the basis of many default color standards. Within the international organization ITU, ITU-R is responsible for the coordination for the efficient use of the radio spectrum and of the geostationary satellite orbit.<sup>2</sup> Within this function, it makes recommendations for television broadcasting systems. The basic colorimetric parameters of Recommendation BT.709 for the HDTV standard are as follows.

- The chromaticity coordinates (x,y) of the primaries are:  
red: (0.640,0.330), green: (0.300, 0.600), blue: (0.150, 0.060).
- The white point is  $D_{65}$ , (x,y) = (0.3127, 0.3290).

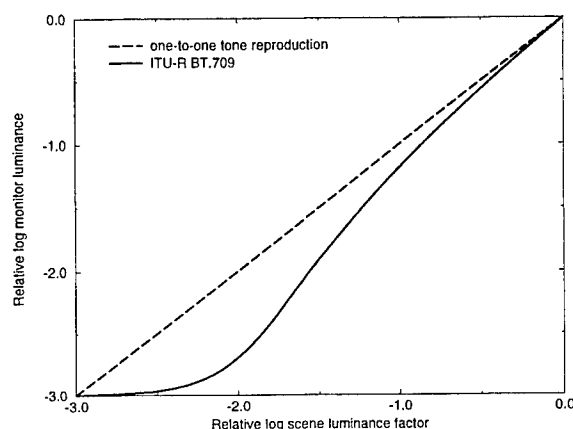
- The overall opto-electronic transfer characteristics at source are:

$$V = 1.099Y^{0.45} - 0.099 \quad \text{for } 0.018 \leq Y \leq 1.0$$

$$V = 4.500Y \quad \text{for } 0.0 \leq Y < 0.018$$

where  $Y$  is the relative luminance of the scene and  $V$  is the corresponding electrical signal.

If we assume that the video signal is displayed on a CRT monitor with a gamma of 2.22 and a viewing flare of 0.1% of the reference white, then the tone reproduction curve for the HDTV images can be derived. The result is shown in Fig. 5. From the figure, it is obvious that the curve has a slope much higher than one, as required



**Figure 5.** The tone reproduction curve used in the HDTV luminance channel as specified by the international standard (ITU-R BT.709).

by the Bartleson-Breneman's brightness model.<sup>46,47</sup> However, if the viewing flare is more than 0.1%, the actual tone reproduction will not have good contrast in the shadow areas.

Recently, sRGB<sup>48</sup> has become a popular default standard color space, which is based on the same primaries and white point as specified in ITU-R BT.709. Since a typical viewing environment of computer monitors is not in a dark surround as was implied by ITU-R BT.709, the sRGB standard changes the reference viewing environment to a dim surround. The sRGB reference viewing environment is assumed to have a 1% veiling flare, an ambient illuminance level of 64 lux with a D<sub>50</sub> ambient illuminant, and a proximal field about 20% of the reflectance of the reference display luminance level, which is at 80 cd/m<sup>2</sup>. These conditions are specified to facilitate the use of color appearance models (such as CIECAM97) for converting one viewing environment to another.

## 4.2. Visual Characterization

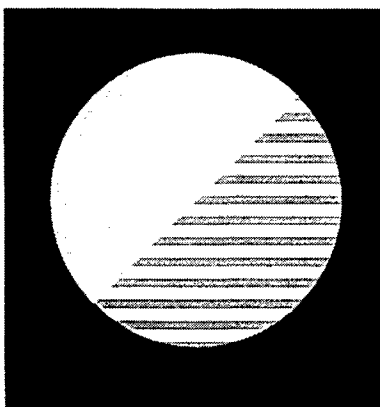
The characteristics of imaging devices change with time. Some devices (such as monitors) allow users to adjust their settings. Therefore, a printer or a monitor might have been well calibrated in the factory, but over its life time, it cannot consistently reproduce colors well without repeated calibration. Use of ICC device profiles or default color spaces cannot solve this type of problem. What is needed for home users is a convenient way to characterize imaging devices. The best solution is for each device to have a built-in internal self-calibration. The next best solution is to have very inexpensive portable instruments to go with an easy-to-use software tool. Since these two solutions tend to increase the cost of the products, a good alternative solution is to use the user's own eyes as an instrument. Test targets can be displayed or printed with well designed patterns. The user can tell the device driver which pattern is best according to the instructed criteria. The driver then uses the user input to select the current, best calibration table for the device. Several such visual characterization methods have been proposed for printers<sup>49-51</sup> and color monitors.<sup>52,50,53</sup>

There are several perceptual phenomena that can be exploited for the visual characterization of displays and printers. The most frequently used one is visual blur. For example, halftone printing using black dots on white paper can generate an image with fine gray scale shadings indistinguishable from a continuous tone image, if the size of the dot is so small as to be blurred together by the optics of our eyes. The following example shows how visual blur can be used to determine the gamma of a CRT monitor.

The basic idea is to model the input-output characteristics of a CRT monitor by a simple equation with a few parameters, and use visual inspection to select the parameter values by choosing the targets that have the right appearances. For example, the luminance as a function of the input digital value of a CRT can be modeled<sup>54,55</sup> as:

$$L = (S - \delta)^\gamma \quad (3)$$

where  $L$  is the relative luminance,  $S$  is the input digital value,  $\delta$  is the offset, and  $\gamma$  is the gamma of the channel being considered. Equation (3) does not take external flare into account, and thus is valid only in a completely darkened room. To simplify the example, we will assume that the offset  $\delta$  has been determined by some other means. We can generate a pattern target that will allow us to determine the correct  $\gamma$  value when it is viewed from a distance. Figure 6 shows a magnified view of the target used for this process. A disk is partitioned into two



**Figure 6.** The disk pattern for estimating the CRT gamma.

halves along a 45-degree line. The upper left half is uniformly filled with a single digital value  $S$ . The lower right half is filled with alternating dark lines and bright lines. The dark lines have a digital value,  $S_1$ , and the bright lines,  $S_2$ . If the user is sufficiently far away from the CRT screen, the dark lines and the bright lines appear to blend together, by the optical blur in the user's eye, to give a shade of gray that is the average luminance of the dark and the bright lines. Two considerations are important for the design of this pattern: (1) The 45 degree boundary is used because our visual system is less sensitive to the oblique direction and therefore can fuse the two sides better when they are of equal luminance. (2) The alternate dark and bright lines, instead of a checker board pattern of dark and bright pixels, are used because most CRTs cannot display on-off patterns fast enough to produce faithful dark and bright pixels.

Given the offset and the gamma, we can calculate the signal value  $S$  on the left half that will match the luminance on the right half:

$$L_1 = (S_1 - \delta)^\gamma \quad (4)$$

$$L_2 = (S_2 - \delta)^\gamma \quad (5)$$

$$L = (S - \delta)^\gamma \quad (6)$$

and

$$L = \frac{1}{2}(L_1 + L_2) \quad (7)$$

$$(S - \delta)^\gamma = \frac{1}{2}[(S_1 - \delta)^\gamma + (S_2 - \delta)^\gamma]. \quad (8)$$

Therefore,

$$S = \left(\frac{1}{2}[(S_1 - \delta)^\gamma + (S_2 - \delta)^\gamma]\right)^{1/\gamma} + \delta \quad (9)$$

To estimate the gamma, we display a series of disks, such as shown in Figure 7, each of which has the same right half with alternating dark and bright lines. The left half is filled with a digital value calculated to match the



**Figure 7.** A series of disks for estimating the CRT gamma.

right half, assuming that the CRT has a certain  $\gamma$  value. For example, the first disk is generated with  $\gamma = 1.5$ , the second with  $\gamma = 1.6$ , the third with  $\gamma = 1.7$ , and so on. If the CRT has a  $\gamma$  value of 2.1, then the disk that was generated with  $\gamma = 2.1$  will look like a uniform disk with both halves appear to have the same luminance. The user's task is to choose, from the array of disks, the one that seems to have the best match of luminances between the left half and the right half. The chosen disk provides the estimate of the CRT  $\gamma$ , i.e.,  $\gamma$  = the gamma value used to generate the selected disk.

A very interesting method conceived by R. L. Gregory for determining the relative "brightness" of different colors is described on page 398 of the book by Kaiser and Boynton.<sup>56</sup> Let a monitor displaying a set of stripes of color A moving to one direction and a set of stripes of color B moving to the opposite direction. Movement is perceived in the direction of the brighter stripes. When both colors are of nearly equal brightness, no drift motion is perceived. Therefore, in principle, it is possible to use this effect to estimate the relative brightness of the red, green, and blue phosphors of a color monitor.

There are many other visual phenomena that are well known in vision research, but have not been well exploited in visual calibration tools. It seems that future research along this direction may produce some solutions to one of the most troublesome problems in Internet color imaging. However, certain visual phenomena are not very sensitive to the variable that we wish to measure. Therefore, search for a robust phenomenon to use is not easy.

## 5. VIEWING CONDITIONS OF UNKNOWN PERCEPTUAL EFFECTS

The environment in which we view an image has very significant effects on our image perception.<sup>46,3,57</sup> There are three major factors to be considered: (1) visual adaptation, (2) surround effect, and (3) viewing flare. Although color appearance models<sup>57</sup> are developed to predict the effects of such factors, they tend to have many parameters that are not easy to adjust for an arbitrary viewing environment. The best solution for this problem is to set up our viewing environment to one of the standard conditions. However, this is not practical in many applications. In terms of what a user can do, reducing flare by turning or shielding the room illumination away from the monitor or viewing a reflection print from an off-specular angle under a directional light source are common sense actions to take.

If we are producing color images that will be viewed under viewing conditions of unknown perceptual effects, the best strategy is to control the dynamic range of the images by spatial processing<sup>58-64</sup> so that details in both



the highlight and the shadow are preserved with good contrast within a compressed luminance dynamic range. Colors need to be made more saturated and white (or gray) borders or backgrounds can be used to help control the chromatic adaptation of the viewer.

## 6. DISCUSSION AND CONCLUSIONS

Standardization across all imaging devices is the main solution to the problems of Internet color imaging. However, standardization does not solve all the problems. The three remaining problems, as discussed in this paper, are quite different in nature and require different types of solution. To deal with color images of unknown calibration, research in computer vision, image understanding, and scene physics will eventually allow us to implement automatic algorithms to handle the problem. To deal with imaging devices of unknown characteristics, inexpensive colorimeters and easy-to-use software calibration tools will be the most feasible solutions in the near future. To deal with viewing conditions of unknown perceptual effects, users can take simple measures to greatly improve their image perception. The alternative solution is to build display devices that can sense the environments and self-adjust their own tone and color reproduction characteristics.

## ACKNOWLEDGMENTS

Many of the results reported here are from the joint work with my colleagues. I would like to thank them and gladly acknowledge their contributions: James Alkofer, John Birkelund, Scott Daly, Robert Goodwin, Heemin Kwon, and Jeanie Liang.

## REFERENCES

1. D. McDowell, "The role and responsibilities of ISO/IEC joint technical advisory group 2 - imagery," *Proc. CIE Expert Symposium '96 on Colour Standards for Image Technology*, pp. 4-6, 1996.
2. C. J. Dalton, K. P. Davies, and O. Gofaizen, "An overview of the activities of the itu in television colorimetry," *Proc. CIE Expert Symposium '96 on Colour Standards for Image Technology*, pp. 18-24, 1996.
3. E. J. Giorgianni and T. E. Madden, *Digital Color Management*, Addison-Wesley, Reading, MA., 1997.
4. M. Stokes and R. Motta, "A default RGB monitor space proposal," *Proc. CIE Expert Symposium '96 on Colour Standards for Image Technology*, pp. 47-50, 1996.
5. M. Stokes, "Color management in the real world: sRGB, ICM2, ICC, ColorSync<sup>TM</sup>, and other attempts to make color management 'transparent'," *Proc. SPIE* **3299**, pp. 360-367, 1998.
6. L. MacDonald, "Colour management and display calibration," *Proc. CIE Expert Symposium '96 on Colour Standards for Image Technology*, pp. 63-69, 1996.
7. A. Johnson, *Colour Management in Graphic Arts and Publishing*, Pira International, Leatherhead, Surrey, UK., 1996.
8. F. Deleixhe-Mauhin, J. M. Krezinski, G. Rorive, and G. E. Pierard, "Quantification of skin color in patients undergoing maintenance hemodialysis," *J. Am. Acad. Dermatol.* **27**, **6**, Part 1, pp. 950-953, 1992.
9. G. E. Pierard, C. Pierard-Franchimont, F. L. Dosal, T. B. Mosbah, J. Arrese-Estrada, A. Rurangirwa, A. Dowlati, and M. Vardar, "Pigmentary changes in skin senescence," *J. Appl. Cosmetol.* **9**, pp. 57-63, 1991.
10. Y. Yamamoto, "Colorimetric evaluation of skin color in the Japanese," *Plast. Reconstr. Surg.* **96**, **1**, pp. 139-145, 1995.
11. E. Edwards and S. Duntley, "Pigment and color in living human skin," *Am. J. Anat.* **65**, pp. 1-33, 1939.
12. E. Edwards and S. Duntley, "Analysis of skin pigment changes after exposure to sunlight," *Science* **90**, pp. 235-237, 1939.
13. G. Burton and I. Moorhead, "Color and spatial structure in natural scenes," *Appl. Opt.* **26**, **1**, pp. 157-170, 1987.
14. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A*, **4**, **12**, pp. 2379-2394, 1987.
15. D. L. Ruderman, "Origins of scaling in natural images," *Vision Res.* **37**, **23**, pp. 3385-3395, 1997.
16. J. S. Alkofer, *Tone Value Sample Selection in Digital Image Processing Method Employing Histogram Normalization*, U.S. Patent No. 4,654,722, Mar. 31, 1987.

17. W. Richards, "Lightness scale from image intensity distribution," *Appl. Opt.* **21**, **14**, pp. 2569–2582, 1982.
18. J. Huang and D. Mumford, "Statistics of natural images and models," *Proc. IEEE Conf. Comput. Vision Pattern Recognit.* **1**, pp. 541–547, 1999.
19. H.-C. Lee and H. Kwon, *Method for Estimating and Adjusting Digital Image Contrast*, U.S. Patent No. 5,822,453, Oct. 13, 1998.
20. L. A. Jones and H. R. Condit, "The brightness scale of exterior scenes and the computation of correct photographic exposure," *J. Opt. Soc. Am.* **31**, **11**, pp. 651–678, 1941.
21. J. S. Alkofer, *Contrast Adjustment in Digital Image Processing Method Employing Histogram Normalization*, U.S. Patent No. 4,731,671, Mar. 15, 1988.
22. R. M. Evans, *Method for Correcting Photographic Color Prints*, U.S. Patent No. 2,571,697, Oct. 16, 1951.
23. J. Hughes and J. K. Bowker, "Automatic color printing techniques," *Image Technol.*, pp. 39–43, April/May 1969.
24. H.-C. Lee, L. L. Barski, and R. A. Senn, *Automatic Tone Scale Adjustment Using Image Activity Measures*, U.S. Patent No. 5,633,511, May 27, 1997.
25. J. R. Boyack and A. K. Juenger, *Brightness Adjustment of Images Using Digital Scene Analysis*, U.S. Patent No. 5,724,456, Mar. 3, 1998.
26. H.-C. Lee and R. Goodwin, "Colors as seen by humans and machines," *Final Program and Advance Printing Papers of the IS&T's 47th Annual Conference*, pp. 401–405, 1994.
27. Y. Satoh, Y. Miyake, H. Yaguchi, and S. Shinohara, "Facial pattern detection and color correction from negative color film," *J. Imaging Technol.* **16**, **2**, pp. 80–84, 1990.
28. D. A. Forsyth and M. M. Fleck, "Automatic detection of human nudes," *Int. J. Comput. Vision* **32**, **11**, pp. 63–77, 1999.
29. M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *Proc. IEEE Conf. Comput. Vision Pattern Recognit.* **1**, pp. 275–281, 1999.
30. M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.* **3**, **1**, pp. 71–86, 1991.
31. G. Yang and T. Huang, "Human face detection in complex background," *Pattern Recognit.* **27**, **1**, pp. 53–63, 1994.
32. K. Yow and R. Cipolla, "Feature-based human face detection," *Image Vision Comput.* **15**, pp. 713–735, 1997.
33. K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, **1**, pp. 39–51, 1998.
34. H. Rowley, S. Baluja, and T. Kanade, "Neural-network-based face detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, **1**, pp. 23–38, 1998.
35. H.-C. Lee, "A physics-based color encoding model for images of natural scenes," *Proc. Conf. Mod. Eng. Technol., Electro-Optics Session*, pp. 25–52, 1992.
36. B. Bayer and P. Powell, "A method for the digital enhancement of unsharp, grainy photographic images," *Adv. Comput. Vision Image Proc.* **2**, pp. 31–88, 1986.
37. D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Am. Statistical Association* **90**, **432**, pp. 1200–1224, 1995.
38. J. Lu, D. M. Healy, Jr., and J. Weaver, "Contrast enhancement of medical images using multiscale edge representation," *Opt. Eng.* **33**, **7**, pp. 2151–2161, 1994.
39. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inf. Theory* **38**, **2**, pp. 587–606, 1992.
40. H.-C. Lee, "A review of of image-blur models in a photographic system using the principles of optics," *Opt. Eng.* **29**, **5**, pp. 405–421, 1990.
41. H.-C. Lee, "Chromatic edge detection: Idealization and reality," *Int. J. Imaging Syst. Technol.* **2**, pp. 251–266, 1990.
42. J. Katajamäki and H. Saarelma, "Objective quality potential measures of natural color images," *J. Imaging Sci. Technol.* **42**, **3**, pp. 250–263, 1998.
43. S. E. Reichenbach, S. K. Park, and R. Narayanswamy, "Characterizing digital image acquisition devices," *Opt. Eng.* **30**, **2**, pp. 170–177, 1991.

44. V. Kayargadde and J. B. Martens, "Estimation of edge parameters and image blur using polynomial transforms," *CVGIP: Graphic Models and Image Processing* **56**, 6, pp. 442-461, 1994.
45. S. I. Olsen, "Estimation of noise in images: an evaluation," *CVGIP: Graphic Models and Image Processing* **55**, 4, pp. 319-323, 1993.
46. A. Johnson and J. Birkenshaw, "The influence of viewing conditions on colour reproduction objectives," *Proc. of the 14th International Conference of Printing Research Institutes*, pp. 48-72, 1977.
47. H.-C. Lee, S. Daly, and R. L. Van Metter, "Visual optimization of radiographic tone scale," *Proc. SPIE* **3036**, pp. 118-129, 1997.
48. *IEC 61966. Part 2.1: Default RGB colour space - sRGB (Third working draft)*, International Electrotechnical Commission, 1998.
49. S. J. Harrington, *Printer Calibration Using a Tone Reproduction Curve and Requiring No Measuring Equipment*, U.S. Patent No. 5,347,369, Sep. 13, 1994.
50. A. D. Edgar and J. M. Kasson, *Display Calibration*, U.S. Patent No. 5,298,993, Mar. 29, 1994.
51. K. A. Hadley and K. E. Spaulding, *Method for Printer Calibration*, U.S. Patent No. 5,995,714, Nov. 30, 1999.
52. R. J. Motta, "Visual characterization of color CRTs," *Proc. SPIE* **1909**, pp. 212-221, 1993.
53. S. J. Daly and H.-C. Lee, *Visual Characterization Using Display Model*, U.S. Patent No. 5,754,222, May 19, 1998.
54. R. Bartow, W. Darrow, and T. Hartmann, *CRT Device Light Versus Input Signal Characteristic Function*, U.S. Patent No. 4,862,265, Aug. 29, 1989.
55. R. Berns, R. Motta, and M. Gorzynski, "CRT colorimetry. Part I: Theory and practice," *Color Res. Appl.* **18**, 5, pp. 299-314, 1993.
56. P. Kaiser and R. Boynton, *Human Color Vision*, Optical Society of America, Washington, D.C., second ed., 1996.
57. M. Fairchild, *Color Appearance Models*, Addison-Wesley, Reading, MA., 1997.
58. E. Wagnsonner, W. Ruf, H. Fuchsberger, and K. Birgmeir, *Method of Electronically Improving the Sharpness and Contrast of a Colored Image for Copying*, U.S. Patent No. 4,812,903, Mar. 14, 1989.
59. H.-C. Lee, M. Kaplan, and R. Goodwin, *An Interactive Dynamic Range Adjustment System for Printing Digital Images*, U.S. Patent No. 5,012,333, Apr. 30, 1991.
60. P. Vuylsteke and E. Schoeters, *Method and Apparatus for Contrast Enhancement*, U.S. Patent No. 5,467,404, Nov. 14, 1995.
61. M. Nakazawa and H. Tsuchino, *Method for Compressing a Dynamic Range for a Radiation Image*, U.S. Patent No. 5,471,987, Dec. 5, 1995.
62. H. Tsuchino and M. Nakazawa, *Radiation Image Processing Method Which Increases and Decreases a Frequency Region of the Radiation Image*, U.S. Patent No. 5,493,622, Feb. 20, 1996.
63. N. Nakajima, *Method for Compressing Dynamic Ranges of Images Using a Monotonously Decreasing Function*, U.S. Patent No. 5,608,813, Mar. 4, 1997.
64. F. Labaere and P. Vuylsteke, *Image Contrast Enhancing Method*, U.S. Patent No. 5,717,791, Feb. 10, 1998.

# Spectral Estimation and Color Appearance Prediction of Fluorescent Materials

Bore-Kuen Lee<sup>a\*</sup>, Feng-Chi Shen<sup>a</sup> and Chun-Yen Chen<sup>b</sup>

<sup>a</sup>Department of Electrical Engineering, Chung-Hua University, Hsinchu City, Taiwan, R.O.C.

<sup>b</sup>Opto-Electronics & Systems Laboratories, Industrial Technology Research Institute, Hsinchu County, Taiwan, R.O.C.

## ABSTRACT

In this paper, we present a method to estimate the reflected and fluorescent spectral radiance factors of a fluorescent object based on spectrophotometric data without using a monochromator. We use truncated Fourier series to approximate both two spectral radiance factors. Then, based on the measured spectral obtained from a spectroradiometer, the coefficients of the truncated Fourier series are estimated using an weighted least square algorithm. The weighting function is defined as the sum of the CIE standard x, y, and z color matching functions. With the estimated reflected and fluorescent spectral radiance factors, we can predict the color appearance of a fluorescent object under other sources such that the color difference is minimized from viewpoint of human vision.

**Keywords:** fluorescent objects, fluorescent spectral radiance factor, spectral estimation, color appearance

## 1. INTRODUCTION

Fluorescent material has been widely applied in various industries. The fluorescent materials have higher lightness and saturation compared with the traditional non-fluorescent materials. The applications of the fluorescent materials include (i) fluorescence dyestuff, such as varnish, toys, printery, and furnishing, (ii) fluorescent ink, such as printing, and warning sign, and (iii) brightening agents such as spin, weave, and papermaking. In industry, the fluorescent materials may be divided into three major categories: brightening agent, the sunlight fluorescent, and inorganic fluorescent.

Though the applications of the fluorescent materials have become more and more important, it is difficult to determine their appearances under different sources. Except for traditional reflected radiance, the fluorescent materials may produce additional radiance called fluorescence, i.e., producing low-frequency radiant energy due to high-frequency energy stimulation. Therefore the chromatic coordinates or the tristimulus values of fluorescent objects will vary dramatically under different light sources. To correctly predict the appearance of a fluorescent object under some specified light source is the main problem for the applications of the fluorescent materials. From the viewpoint of spectrophotometry<sup>1</sup>, the fluorescence phenomenon can be characterized by the sum of the one-dimensional reflected spectral radiance factor  $\beta_s$  and the two-dimensional fluorescent spectral radiance factor  $\beta_L$ . The key issue to study a fluorescent material is to determine these two spectral radiance factor functions.

Several methods<sup>1,2,6</sup> had been proposed to determine the spectral reflected radiance factor  $\beta_s$  and spectral luminescent radiance factor  $\beta_L$ . For the one-monochromator method<sup>1</sup>, light irradiates on the fluorescent materials and a monochromator is used to control the wavelength range measured by detectors. By using the one-monochromator method, we can obtain the overall spectral radiance factor  $\beta_T$ , which is the sum of the spectral reflected radiance factor  $\beta_s$  and spectral luminescent radiance factor  $\beta_L$ . However, the spectral reflected radiance factor  $\beta_s$  and spectral luminescent radiance factor  $\beta_L$  can not be distinguished from  $\beta_T$ . The experiment of the one-monochromator method is easy to establish. However, as the fluorescent spectral radiance function  $\beta_L$  can not be obtained, this method can be applied to determine the color appearance of the fluorescent object under a specified source.

\* Correspondence: Email: bklee@chu.edu.tw; Telephone: +886-3-5374281 ext. 8511; Fax: +886-3-5374281 ext. 8930

For the two-monochromator method<sup>1,2,3,4,5</sup>, two monochromators are used as a monochromatic radiation selector and a radiant detector, respectively. When light passes through the first monochromator, the monochromatic radiation at selected wavelength irradiates the fluorescent material. The second monochromator then measure the reflected and re-emitted radiation from the fluorescent material at various wavelengths. Therefore we can obtain a two-dimensional reflected matrix representing the spectrophotometric property of the fluorescent material. The spectral reflected radiance factor  $\beta_s$  is usually determined from the diagonal of the matrix and the spectral luminescent radiance factor  $\beta_L$  from the off-diagonal part. However, significant error for the determined  $\beta_s$  may happen at the wavelengths in the intersection of the excitation band and the emission band of the fluorescent material.

With the two-mode method<sup>2</sup>, we can obtain the overall spectral radiance factor and spectral conventional reflectometer value by white irradiating and monochromatic irradiating. Therefore, the overall spectral radiance factor will equal to spectral reflected radiance factor at short-wavelengths. And the conventional reflectometer reading is as the same as the spectral reflected factor at long-wavelengths. Within excitation band and emission band of the fluorescent material, interpolation is used to estimate the spectral reflected radiance factor. Hence, the spectral reflected radiance factor can be determined. And the spectral luminescent radiance factor can be calculated as the difference between the overall spectral radiance factor and the spectral reflected radiance factor. Though, we can determined the spectral reflected radiance, it is difficult to obtain a good estimated spectral reflected radiance factor in both excitation band and emission band of the fluorescent material.

The filter reduction method and the luminescence-weakening method were further proposed<sup>2,6</sup> to provide a better method to estimated the spectral reflected radiance factor in fluorescent excitation and emission region. With the filter reduction method, one can obtain the overall spectral radiance factor by using white light to irradiate the sample. Therefore the overall spectral radiance factor will be equal to the spectral reflected radiance factor at short wavelengths. Then the spectral reflected radiance factor at long wavelengths will be obtained by placing a sharp cut-off filter between the source and the sample. Within excitation band and emission band of the fluorescent material, interpolation is used to estimate the spectral reflected radiance factor. A series of filters are used to reduce the fluorescent emission and to derive an approximation of the spectral fluorescent radiance factor. Because the filter reduction method inducted a series sharp cut-off filter to give an approximation of the spectral reflected radiance factor within excitation band and emission band of the fluorescent material, this method improves estimation accuracy of the spectral fluorescent radiance factor. When samples have a small amount of fluorescence, it is better to use the filter-reduction method. On the contrary, it is preferable to use the luminescence-weakening method<sup>2</sup> in case of large amount of fluorescence. In the luminescence-weakening method, we can obtain the overall spectral radiance factor by using the white light to irradiate the sample. Therefore, the overall spectral radiance factor will equal to spectral reflected radiance factor at short-wavelengths. Then the spectral reflected factor at long-wavelengths will be obtained by placing a complete cut-off filter between the source and the sample. Within excitation band and emission band of fluorescent material, the spectral reflected radiance factor will be deduced by inducting a partly filter just above the minimum of the overall spectral radiance factor. The particular wavelength must choose carefully, otherwise it will induce significant error.

Motta and Farrell<sup>7</sup> used sixteen filtered tungsten lights to generate sixteen sources for which their spectra can be used to approximate most illuminates. Then a matrix is established to present the radiance of the fluorescent material under the sixteen sources. If the vector representing the characteristics based on the sixteen generated sources for a given source can be obtained, then the colorimetric information of the fluorescent material under such a source can be estimated. However, these estimated colorimetric data are not optimized from human vision.

Based on the previous discussions, we now summarize as follows. For the the two-monochromator method and its varieties such as the two-mode method, the filter reduction method, and the luminescence-weakening method, the reflected spectral radiance factor  $\beta_s$  in the middle wavelengths is estimated by interpolating the spectral radiance data of the short and long wavelengths. This rough interpolation cause estimation error of  $\beta_s$  and thus of  $\beta_L$ . Next, the monochromator is used in all the work discussed above. Monochromator is expensive and hard to maintain and operate. The most important thing, which is neglected, is that we should further validate the measured or estimated spectral radiance factors, for example, by predicting the color appearance under some given source. In this paper, we shall construct a method to estimate the reflected and fluorescent spectral radiance factor of a fluorescent object based on spectrophotometric data without using a monochromator. We shall use truncated Fourier series to approximate both two spectral radiance factors. Then, based on the measured spectral distribution, the coefficients of the truncated Fourier series will be estimated using an weighted least square algorithm. The weighting function is constructed by using the CIE standard color matching functions. Then we shall provide an estimated appearance of a fluorescent object under given sources such that the color difference is minimized from viewpoint of human vision.

The remaining of this paper is organized as follows. The spectral properties of fluorescent materials are discussed in Section 2. In Section 3, we shall describe the setup of our experimental environment. A least algorithm, which is used to estimate the spectral reflected radiance factor and the spectral fluorescent radiance factor, and experimental results are described in Section 4. Finally, conclusions are given in Section 5.

## 2. SPECTROPHOTOMETRIC ANALYSIS AND COLORIMETRIC MEASUREMENT OF FLUORESCENT MATERIALS

Usually, a fluorescent material is made of a substrate by adding some fluorescent agents. Therefore, due to fluorescence, the overall spectral radiance factor  $\beta_T(\lambda, u)$  of a material consists of two parts:

$$\beta_T(\lambda, u) = \beta_S(\lambda) + \beta_L(\lambda, u) \quad (1)$$

where  $\beta_S(\lambda)$  is the spectral radiance factor reflected by the substrate and  $\beta_L(\lambda, u)$  is that accounts for the fluorescence effect of the material. From viewpoint of spectrophotometric analysis made in equation (1), to determine the appearance of a fluorescent material irradiated by different source, it is important to find the two-dimensional spectral radiance function  $\beta_L(\lambda, u)$ .

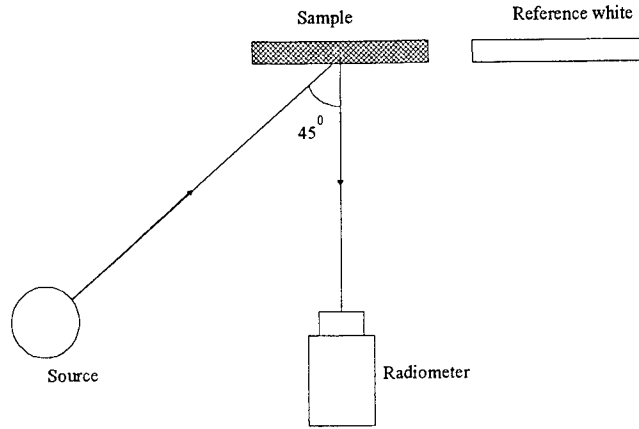


Figure 1: The colorimetric measurement setup

Now we turn to discuss colorimetric measurement of fluorescent materials under the measurement structure shown in Figure 1. Let  $S(\lambda)$ ,  $L_{prd}(\lambda)$ ,  $L_{obj\_flu}(\lambda, u)$ , and  $L_{obj\_sub}(\lambda)$  be the spectral irradiance of the source, the spectral radiance of the reference white, the spectral radiance of the fluorescent agents, and the spectral radiance of the substrate, respectively. Then we define the spectral radiance coefficients of the reference white and the agents of fluorescent sample as

$$q_{prd}(\lambda) = \frac{L_{prd}(\lambda)}{S(\lambda)} \quad (2)$$

$$q_{obj}(\lambda, u) = \frac{L_{obj\_flu}(\lambda, u)}{S(\lambda)} \quad (3)$$

where  $u$  is the excited wavelength and  $\lambda$  is the emitted wavelength. Next, the spectral radiance factor of the fluorescent agent is defined as

$$\beta_L(\lambda, u) = \frac{q_{obj\_flu}(\lambda, u)}{q_{prd}(\lambda)} \quad (4)$$

With the definitions made in equations (2) and (3), the spectral function  $\beta_L(\lambda, u)$  in (4) can be rewritten as

$$\beta_L(\lambda, u) = \frac{L_{obj\_flu}(\lambda, u)}{L_{prd}(\lambda)} \quad (5)$$

Similarly, the spectral radiance factor of the substrate of fluorescent material is defined as

$$\beta_S(\lambda) = \frac{L_{obj\_sub}(\lambda)}{L_{prd}(\lambda)} \quad (6)$$

Then the overall spectral distribution  $C(\lambda)$  measured by the radiometer for the fluorescent sample is given by

$$C(\lambda) = \int_u \beta_L(\lambda, u) S(u) du + \beta_S(\lambda) S(\lambda) \quad (7)$$

By using the spectral function  $C(\lambda)$ , the colorimetric tristimulus values of the fluorescent under light source can be calculated as

$$X_T = 100 \frac{\int_{\lambda} C(\lambda) \bar{x}(\lambda) d\lambda}{\int_{\lambda} S(\lambda) \bar{y}(\lambda) d\lambda} \quad Y_T = 100 \frac{\int_{\lambda} C(\lambda) \bar{y}(\lambda) d\lambda}{\int_{\lambda} S(\lambda) \bar{y}(\lambda) d\lambda} \quad Z_T = 100 \frac{\int_{\lambda} C(\lambda) \bar{z}(\lambda) d\lambda}{\int_{\lambda} S(\lambda) \bar{y}(\lambda) d\lambda}$$

where  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  are the CIE 1931 standard color matching functions.

### 3. EXPERIMENT SETUP

#### 3.1 Setup of experimental environment

In order to establish colorimetric and radiometric measurements for the fluorescent samples, we set up an experiment as shown in Figure 1. The measurement instrument setup follows the 45/0 geometry. Both the fluorescent sample and the reference white are irradiated by a source and their spectral radiance are measured by a spectroradiometer.

#### 3.2 Selection of Light Sources

To select the sources, we must consider the fluorescence phenomena of the samples. First, the fluorescence is excited within the short-wavelength range of the incident radiant energy. Next, the intensity of the fluorescent spectral radiance factor  $\beta_L(\lambda, u)$  is less than that of the reflected spectral radiance factor  $\beta_S(\lambda)$ . Therefore a source with its power distributed over long-wavelength range can be used to effectively estimate the spectral function  $\beta_S(\lambda)$ . However, to measure the fluorescent spectral function  $\beta_L(\lambda, u)$ , we need to select a source with rich excitation over short-wavelength range. We use seven different sources in our experiment. These sources are:

- Source 1: a Tungsten lamp,
- Source 2: a test lamp from CVI corporation,
- Source 3: a Xenon lamp,
- Source 4: an approximate of the illuminant  $D_{65}$ ,
- Source 5: the first test fluorescent lamp,
- Source 6: the second test fluorescent lamp,
- Source 7: the third test fluorescent lamp.

The relative spectral power distribution are show in Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Figure 7, and Figure 8, respectively. Note that the power spectrums of Source 1 and Source 2 are mainly distributed over long-wavelength range while the excitation of short-wavelength range is richer for Source 3, Source 4, Source 5, Source 6, and Source 7.

#### 3.3 Descriptions of the Fluorescent Samples

We will estimate the color appearance of three kinds of fluorescent samples under different sources. There are three fluorescent samples, which are yellow, pink, and white, respectively, to be tested.

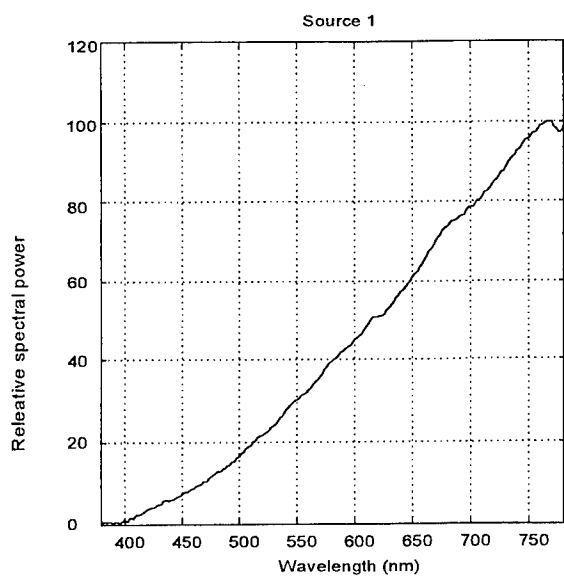


Figure 2: Relative spectral power distribution of Source 1 with 2nm sampling between 380-780nm

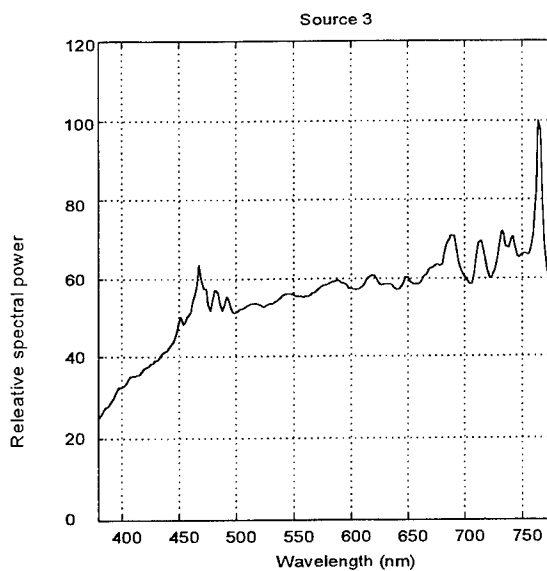


Figure 4: Relative spectral power distribution of Source 3 with 2nm sampling between 380-780nm

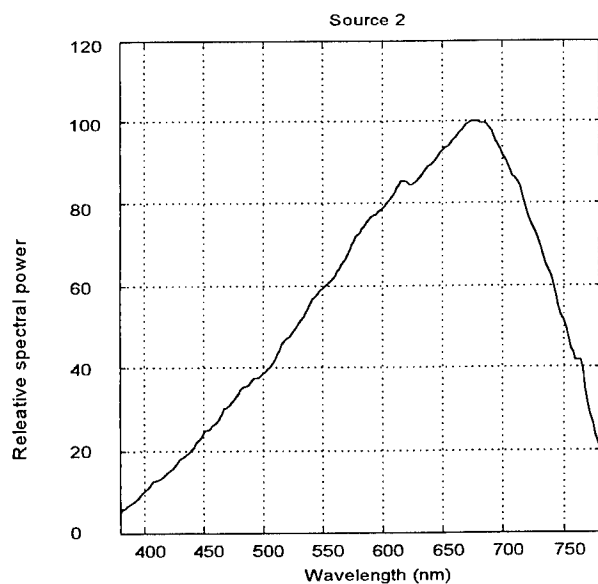


Figure 3: Relative spectral power distribution of Source 2 with 2nm sampling between 380-780nm

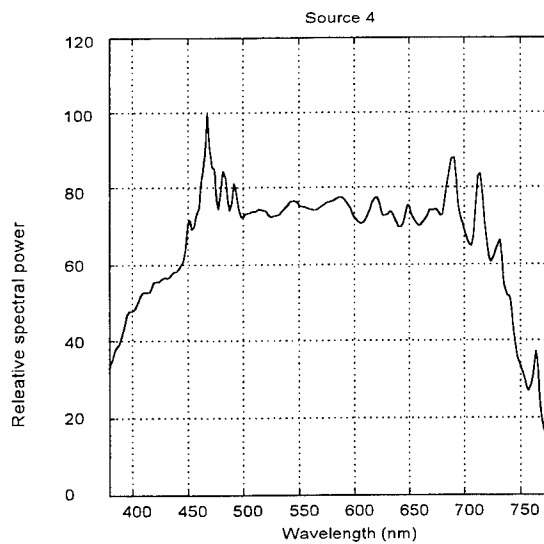


Figure 5: Relative spectral power distribution of Source 4 with 2nm sampling between 380-780nm



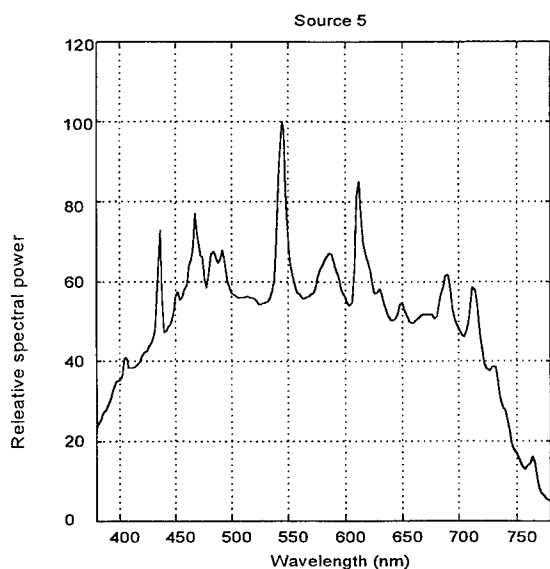


Figure 6: Relative spectral power distribution of Source 5 with 2nm sampling between 380-780nm

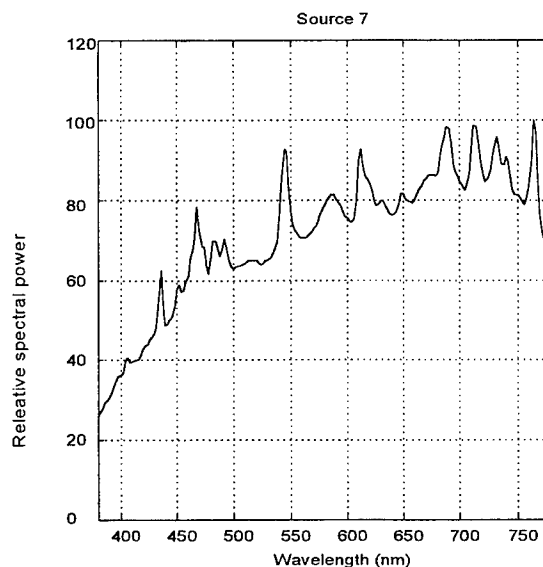


Figure 8: Relative spectral power distribution of Source 7 with 2nm sampling between 380-780nm

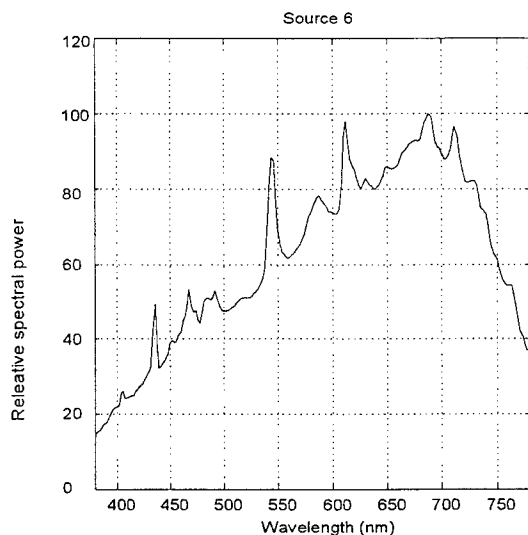


Figure 7: Relative spectral power distribution of Source 6 with 2nm sampling between 380-780nm

### 3.4 The Observer

The definition of observer in our experiment follows the CIE 1931 Standard Observer. The purpose of using the 45/0 geometry is to avoid re-excitation of the fluorescent materials. We use a spectroradiometer (PR-704) to measure the spectral distribution of the sample. The spectral distribution is sampled every 2nm within the wavelength range from 380nm to 780nm.

### 3.5 Measurement Procedure

The spectral radiance of the reference white irradiated by the three different sources, Source 1 to Source 3, are measured by the spectroradiometer. Then, the spectral radiance function  $C(\lambda)$  for each tested fluorescent sample is measured with respect

to the three light sources. In order to reduce the effect of random fluctuation<sup>1</sup>, each spectral distribution is obtained by averaging ten measurements.

#### 4. SPECTRAL ESTIMATION ALGORITHM AND EXPERIMENTAL RESULTS

To estimate the spectral functions  $\beta_s(\lambda)$  and  $\beta_L(\lambda, u)$ , we use Fourier series expansions to represent these two functions. and equation (7) is transformed into the linear regression form in order to apply standard least square algorithm. In particular, to minimize the estimation error of the color appearance of a fluorescent sample, the cost function involved in the least square algorithm is weighted by a function which is the summation of the CIE color-matching functions  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$ . For each fluorescent sample, the measured spectral distribution under Source 1, Source 2, and Source 3, are used in the least squared algorithm to obtain the estimated functions  $\hat{\beta}_s(\lambda)$  and  $\hat{\beta}_L(\lambda, u)$ . These estimated spectral radiance functions are then used to predict the tristimulus of the tested fluorescent samples under Source 4, Source 5, Source 6, and Source 7.

##### 4.1 Least Square Estimation Algorithm

For representing the spectral function  $\beta_L(\lambda, u)$ , we use a truncated two-dimension Fourier sine series expansion; and for the spectral reflected radiance factor  $\beta_s(\lambda)$ , a truncated one-dimension Fourier sine series expansion is used as follows. The estimated fluorescent spectral radiance factor is of the form:

$$\hat{\beta}_L(\lambda, u) = \sum_{n=1}^{N-1} \sum_{m=1}^{N-1} \theta_{nm} \sin\left(\frac{n(\lambda - \lambda_0)\pi}{K}\right) \sin\left(\frac{m(u - u_0)\pi}{L}\right) \quad (8)$$

where  $K=L=400\text{nm}$ ,  $\lambda_0 = u_0 = 380\text{nm}$ , and  $\theta_{nm}$  are the coefficients to be determined. The constant  $N=26$  is the number of sinusoidal function used in the Fourier series expansion with respect to both the  $u$  axis and the  $\lambda$  axis. The structure of the estimated reflected spectral radiance function  $\hat{\beta}_s(\lambda)$  is given by

$$\hat{\beta}_s(\lambda) = \sum_{\alpha=1}^{\mathcal{G}-1} \delta_{\alpha} \sin\left(\frac{\alpha(\lambda - \lambda_0)\pi}{\Phi}\right) + \delta_0 \quad (9)$$

where  $\Phi=400\text{nm}$  and  $\mathcal{G}=51$  is the number of coefficients,  $\delta_{\alpha}$  and  $\delta_0$ , in this series to be determined. Note that the number of all the coefficients to be determined in (8) and (9) is 676. This number is much less than that required in the two-monochromator method.

Since the spectral distribution  $C(\lambda)$  is measured every 2nm, the estimated total spectral radiance distribution  $\hat{C}(\lambda)$  is defined as

$$\hat{C}(\lambda) = \sum_{\substack{u=380 \\ u \text{ even}}}^{780} \sum_{n=1}^{N-1} \sum_{m=1}^{N-1} \theta_{nm} \sin\left(\frac{n(\lambda - \lambda_0)\pi}{K}\right) \sin\left(\frac{m(u - u_0)\pi}{L}\right) S(u) \Delta u + \left( \sum_{\alpha=1}^{\mathcal{G}-1} \delta_{\alpha} \sin\left(\frac{\alpha(\lambda - \lambda_0)\pi}{\Phi}\right) + \delta_0 \right) S(\lambda) \quad (10)$$

where  $\Delta u = 2 \text{ nm}$ . The above expression can be rewritten into linear regression which fits the setup in least square estimation. Define a regression vector  $h(\lambda)$  and a parameter vector  $\phi$  as

$$\begin{aligned} h(\lambda) &= [f_{1,1}(\lambda) \cdots f_{1,N-1}(\lambda) f_{2,1}(\lambda) \cdots f_{N-1,N-1}(\lambda) r_1(\lambda) \cdots r_{\mathcal{G}-1}(\lambda) r_0(\lambda)]^T \\ \phi &= [\theta_{1,1} \cdots \theta_{1,N-1} \theta_{2,1} \cdots \theta_{N-1,N-1} \delta_1 \cdots \delta_{\mathcal{G}-1} \delta_0]^T \end{aligned}$$

where the superscript  $T$  denotes matrix transpose,  $r_0(\lambda) = S(\lambda)$ , and

$$\begin{aligned} f_{nm}(\lambda) &= \sin\left(\frac{n(\lambda - \lambda_0)\pi}{K}\right) \left( \sum_{\substack{u=380 \\ u \text{ even}}}^{780} \sin\left(\frac{m(u - u_0)\pi}{L}\right) S(u) \Delta u \right), \quad n, m = 1, \dots, N-1 \\ r_{\alpha}(\lambda) &= \sin\left(\frac{\alpha(\lambda - \lambda_0)\pi}{\Phi}\right) S(\lambda), \quad \alpha = 1, \dots, \mathcal{G}-1 \end{aligned}$$

Then equation (10) can be rewritten into a linear regression form as

$$\hat{C}(\lambda) = h^T(\lambda)\phi \quad (11)$$

For any given estimated vector  $\phi$ , the spectral estimated error at wavelength  $\lambda$  is given by  $C(\lambda) - h^T(\lambda)\phi$ . Let  $\xi(\lambda)$  be a weighting function defined as

$$\xi(\lambda) = \bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda) \quad (12)$$

where  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  are the CIE 1931 standard color-matching functions. The weighting function  $\xi(\lambda)$  is used to emphasize the effective wavelength range for human vision. By the definitions of the CIE 1931 standard color-matching functions, we can find that the effective wavelength region for human vision is between 380nm and 750nm which is the union of the supports of the functions  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$ . In order to minimize the color difference from human vision coordinate (X, Y, Z), the weighting function  $\xi(\lambda)$  is included in the least square algorithm to specify the wavelength range within which the spectral estimation error is more significant.

There are three sets of spectral distributions irradiated by Source 1 to Source 3 to be used for training the estimated parameter vector  $\hat{\phi}$ . Each spectral set is sampled every 2nm from 380nm to 780nm. Denote  $C_j(\lambda)$ ,  $j=1,2,3$ , as the measured overall spectral radiance function and  $h_j(\lambda)$  as its related regression vector under Source 1, Source 2, and Source 3, respectively. Define a sequence of wavelengths as  $\lambda_j = 380 + 2j$  nm,  $j = 0, \dots, 200$ . Also define the following vectors and matrix as

$$\begin{aligned} \bar{\xi}^T &= [\xi(\lambda_0) \dots \xi(\lambda_{200}) \quad \xi(\lambda_0) \dots \xi(\lambda_{200}) \quad \xi(\lambda_0) \dots \xi(\lambda_{200})] \\ \bar{C}^T &= [C_1(\lambda_0) \dots C_1(\lambda_{200}) \quad C_2(\lambda_0) \dots C_2(\lambda_{200}) \quad C_3(\lambda_0) \dots C_3(\lambda_{200})] \\ \bar{h}^T &= [h_1(\lambda_0) \dots h_1(\lambda_{200}) \quad h_2(\lambda_0) \dots h_2(\lambda_{200}) \quad h_3(\lambda_0) \dots h_3(\lambda_{200})] \end{aligned}$$

Now denote  $\bar{\xi}_i$  and  $\bar{C}_i$  as the  $i$ -th entry of the vectors  $\bar{\xi}$  and  $\bar{C}$ , respectively. Let  $\bar{h}_i$  be the transpose of the  $i$ -th row vector of  $\bar{h}$ . Then the cost function to be minimized in the least square algorithm is defined as

$$J(\phi) = \frac{1}{2} \sum_{i=1}^{603} \bar{\xi}_i (\bar{C}_i - \bar{h}_i^T \phi)^2 \quad (13)$$

The best parameter vector estimate which minimizes the cost function  $J(\phi)$  is denoted as  $\hat{\phi}_{LS}$ , i.e.,

$$\hat{\phi}_{LS} = \arg \min_{\phi} J(\phi)$$

A recursive algorithm to compute  $\hat{\phi}_{LS}$  is given by

$$\begin{aligned} \hat{\phi}_i &= \hat{\phi}_{i-1} + \frac{P_{i-2} \bar{\xi}_{i-1} \bar{h}_{i-1}}{1 + \bar{h}_{i-1}^T P_{i-2} \bar{h}_{i-1}} [\bar{C}_i - \bar{h}_{i-1}^T \hat{\phi}_{i-1}] \\ P_{i-1} &= P_{i-2} - \frac{P_{i-2} \bar{h}_{i-1} \bar{h}_{i-1}^T P_{i-2}}{1 + \bar{h}_{i-1}^T P_{i-2} \bar{h}_{i-1}} \end{aligned} \quad i = 2, \dots, 603 \quad (14)$$

where the initial data  $\hat{\phi}_1$  can be arbitrarily given and  $P_0$  must be a positive definite matrix. The last estimate  $\hat{\phi}_{603}$  is set to  $\hat{\phi}_{LS}$ .

By the proposed weighted least square algorithm, we can obtain the estimated reflected spectral radiance function  $\hat{\beta}_s(\lambda)$  and the estimated fluorescent spectral radiance function  $\hat{\beta}_L(\lambda, u)$  for each fluorescent sample. By using equation (10), we can construct the spectral radiance function  $\hat{C}(\lambda)$  that estimates the spectral radiance function  $C(\lambda)$  for the test fluorescent samples under Source 4, Source 5, Source 6, and Source 7. Comparisons of the estimated overall spectral radiance function

$\hat{C}(\lambda)$  and the measured one  $C(\lambda)$  for the three fluorescent samples under different sources are illustrated in Figure 9 to Figure 14. In these figures, the measured spectral function  $C(\lambda)$  is well approximated by its estimated one. The approximation error for the measured spectral radiance function  $C(\lambda)$  leads to color difference. Color differences in terms of CIELAB Color Difference Formulae for each fluorescent sample under Source 1 to Source 7 are presented in Table 1. The values of  $\Delta E_{ab}^*$  under Source 1, Source 2, and Source 3 in Table 1 are very small because the measured spectral distribution under these sources are used to train the parameter estimates in the least square algorithm. The estimated spectral functions  $\hat{\beta}_s(\lambda)$  and  $\hat{\beta}_L(\lambda, u)$  are then used to predict the color appearance of fluorescent samples under Source 4, Source 5, Source 6, and Source 7. Since the measured spectral data under Source 4, Source 5, Source 6, and Source 7 are not used in the least square algorithm, the values of  $\Delta E_{ab}^*$  under these sources are usually larger. It is more difficult to decrease the values of  $\Delta E_{ab}^*$  under Source 5, Source 6, and Source 7 since these sources are mainly based on fluorescent lamps, which have discrete spectral distributions.

$\Delta E_{ab}^*$	Source 1	Source 2	Source 3	Source 4	Source 5-7
Sample 1	0.0010	0.0001	0.0011	3.55	6.73(under Source 5)
Sample 2	0.0013	0.00001	0.0014	3.59	8.81(under Source 6)
Sample 3	0.0130	0.0060	0.0003	4.6	4.41(under Source 7)

Table 1: Color differences of the fluorescent samples under different sources

## 5. CONCLUSIONS

In this paper, we have presented a method to estimate the reflected and fluorescent spectral radiance factor of a fluorescent object based on spectrophotometric data without using a monochromator. We use truncated Fourier series to approximate both two spectral radiance factors. The total number of parameters to be determined and recorded are much less than that in the two-monochromator method. With the special weighted least square algorithm, we are able to provide estimated color appearance of test fluorescent objects under different sources such that the color difference is minimized from viewpoint of human vision. With the experiment test made in Section 4, the proposed method can be effectively used to predict color appearance for the applications of fluorescent materials such as fluorescent dyestuff, fluorescent ink, and brightening agents.

## ACKNOWLEDGMENTS

This study and related instruments are supported by Opto-Electronics & Systems Laboratories of Industrial Technology Research Institute.

## REFERENCES

1. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2<sup>nd</sup> edition., John Wiley & Sons, New York, 1982.
2. R. W. G. Hunt, *Measuring colour*, Ellis Horwood, New York, 1991.
3. D. Gundlach and H. Terstiege, "Problems in Measurement of Fluorescent Materials," *Color Research and Application*, Vol 19, Num. 6, December, 1994.
4. Frederick T. Simon, Rober A. Funk and Ann Campbell Laidlaw, "Match Prediction of Highly Fluorescent Colors," *Color Research and Application*, Vol 19, Num 6, December, 1994.
5. *Methods of Measurement for colour of Fluorescent Object*, JIS Z 8717.

6. F. W. Billmeyer, "Metrology, Documentary Standards, and Color Specifications for Fluorescent Materials," *Color Research and Application*, Vol 19, Num 6, December, 1994.
7. Ricardo Motta, and Joyce Farrel, "A simplified Method for Colorimetric Characterization of Fluorescent Inks," *IS&T and SID's Color Imaging Conference: Transforms & Transportability of Color*, 1993.
8. L. Ljung, *System Identification: Theory for the User*, Prentice Hall, New York, 1987.

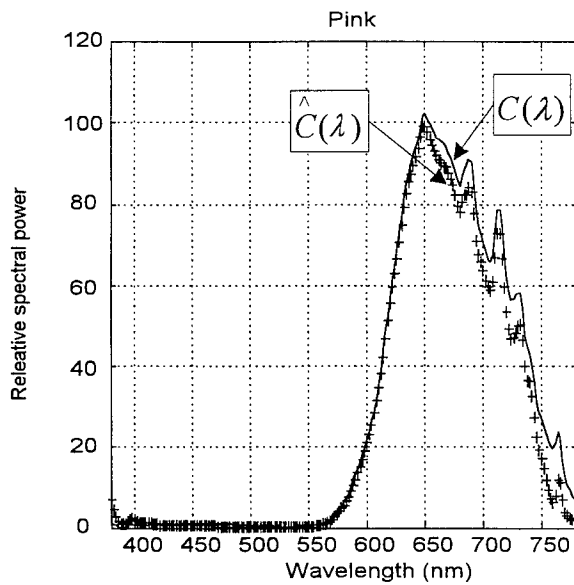
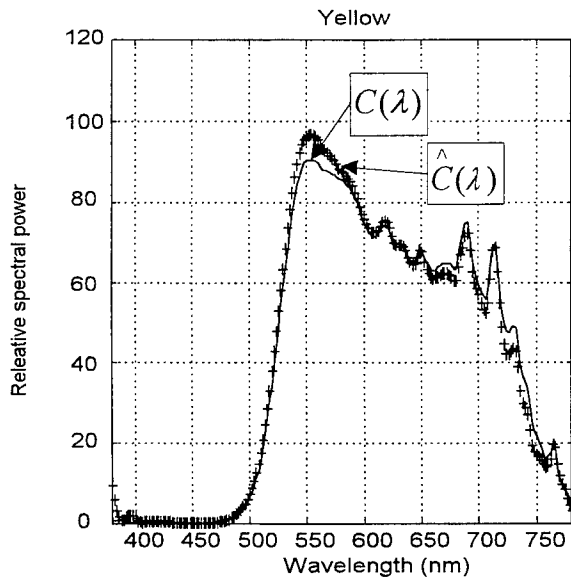
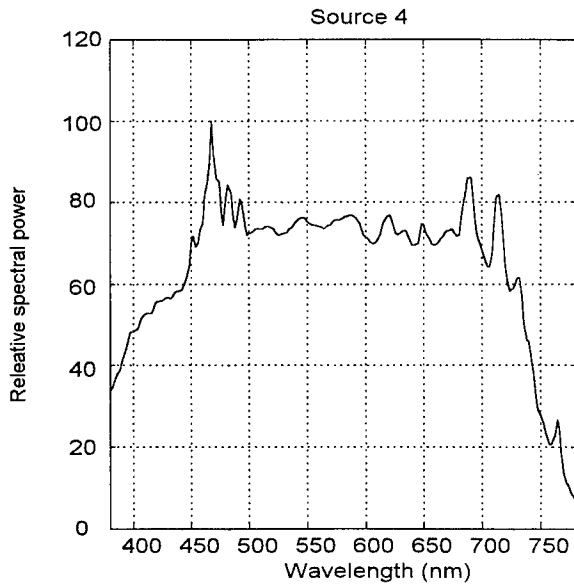
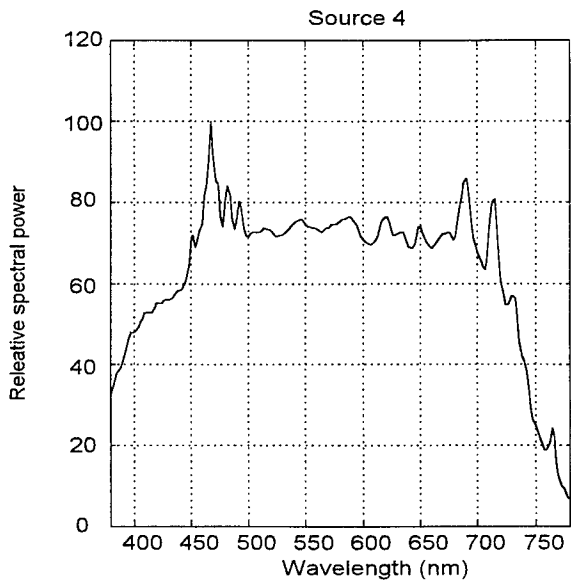


Figure 9: Comparison of  $C(\lambda)$  and  $\hat{C}(\lambda)$  for the yellow fluorescent sample under Source 4.

Figure 10: Comparison of  $C(\lambda)$  and  $\hat{C}(\lambda)$  for the pink fluorescent sample under Source 4.

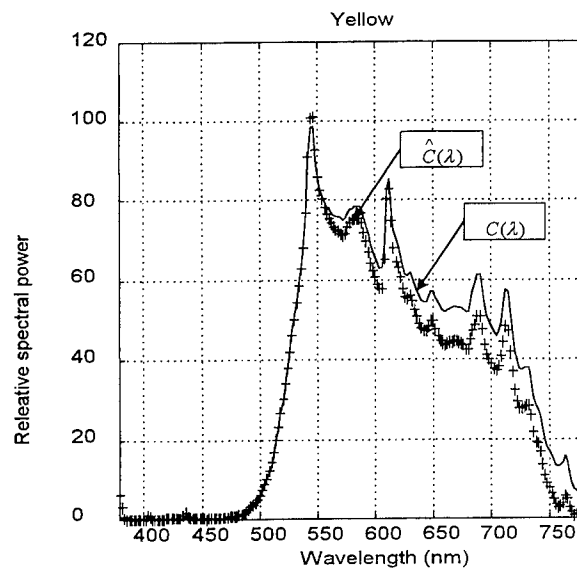
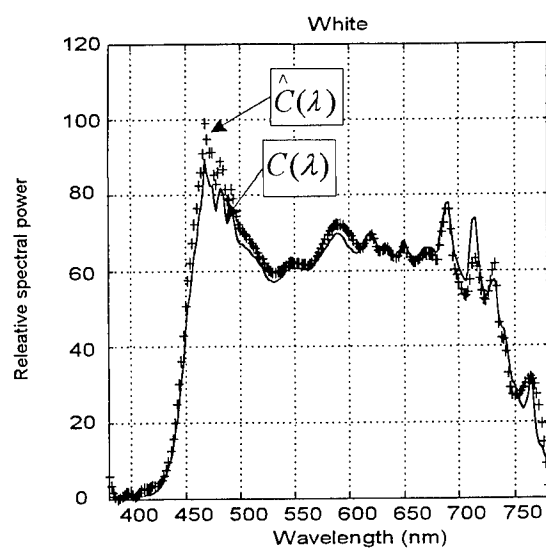
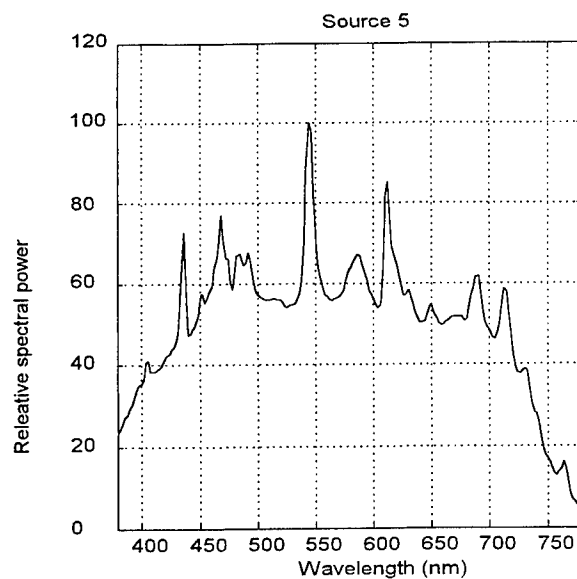
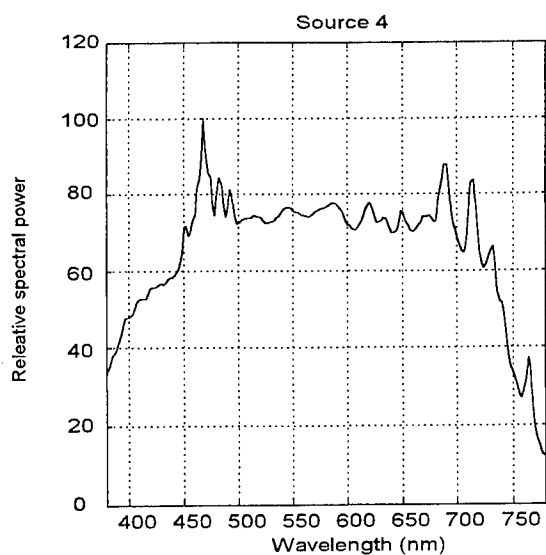


Figure 11: Comparison of  $C(\lambda)$  and  $\hat{C}(\lambda)$  for the white fluorescent sample under Source 4.

Figure 12: Comparison of  $C(\lambda)$  and  $\hat{C}(\lambda)$  for the yellow fluorescent sample under Source 5.

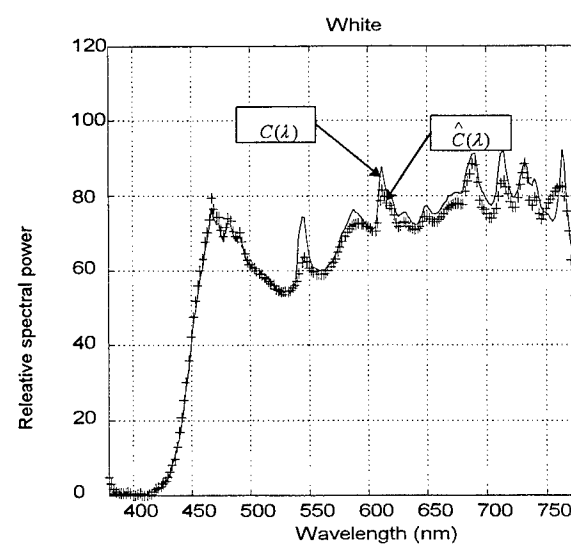
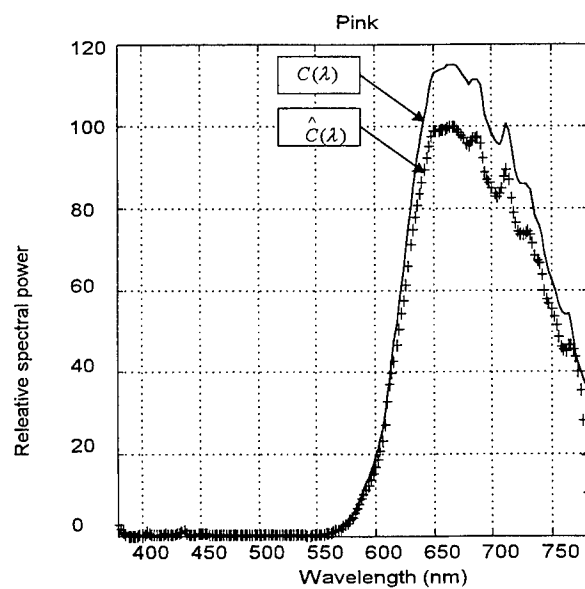
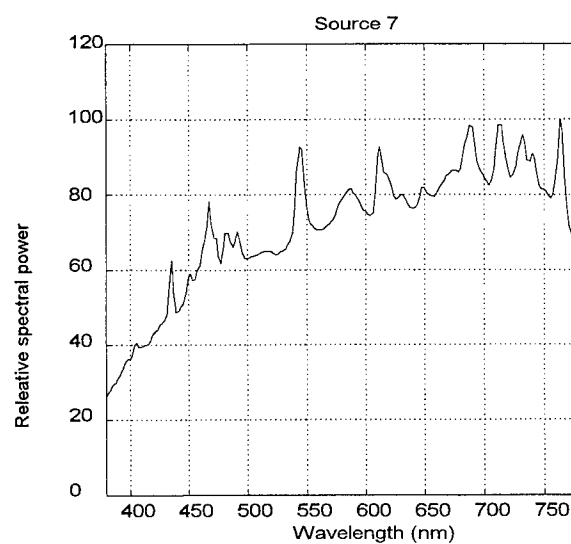
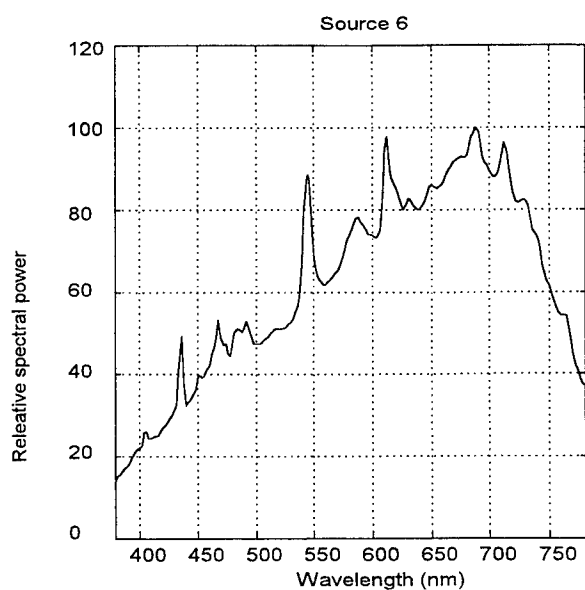


Figure 13: Comparison of  $C(\lambda)$  and  $\hat{C}(\lambda)$  for the pink fluorescent sample under Source 6.

Figure 14: Comparison of  $C(\lambda)$  and  $\hat{C}(\lambda)$  for the white fluorescent sample under Source 7.

# Design and production of color calibration targets for digital input devices

Chao-hua Wen and Jyh-jiun Lee

Opto-Electronics & Systems Laboratories, Industrial Technology Research Institute  
Hsin-chu, Taiwan, ROC

## ABSTRACT

This paper presents the design and production of calibration targets for digital input color devices. By experimentally determined gamut of surface color, this study redesigns the aim values based on ISO/FDIS 12641 and to meet process specifications of Noritsu QSS23-HRCRT photographic printer with silver halide photography. The calibration target includes four components: a set of 144 color patches (3 levels in lightness and 4 levels in chroma at 12 different hue angles) within printing gamut, a neutral scale containing 22 steps based on visual perception, a set of C-M-Y-K-R-G-B dye scales showing characteristics of photographic materials, and a series of facial colors ranked by red. This research will describe the meaning of each element, the use of colorimetric mapping to CIELCH for each element, the conversion of these patch into a RGB-mode electronic image file, and how to control the processing of color photographic materials. And we propose an approach of dynamic subgroup linear interpolation to achieve high process quality of manufacturing calibration targets and cost-down. Finally, statistic results revealed that 99% of the patches are within 10 delta Eab of the aim values specified in this study from long-term test and 99% of the patches in the manufacturing batch are within 5 delta Eab of the mean values from short-term test.

**Keywords:** Calibration target, Test chart, Color gamut, Process control

## 1. INTRODUCTION

While color digital input devices are now used to capture images for output on a variety of media, their development has essentially taken place within a Graphic Arts environment. Digital input devices have the same problem that is the task of making a colored image in a reproduction "match" a colored image in the original under some specified illuminant in Graphic Arts. The most efficient method for characterizing a scanner or digital camera is to image a set of aimed colors of known tristimulus values.<sup>1</sup> Using color calibration techniques, the data obtained from this imaging process can be compared to the tristimulus values of the test image and a colour transformation defined.

A range of tools has been developed by ANSI, consisting of a series of photographic prints and transparencies, which are used for characterizing an input scanner. The specifications are now being circulated by ISO as ISO 12641.<sup>2</sup> A color test image has been designed based on the Kodak Q60<sup>TM</sup> test transparency described by Maier and Rinehart,<sup>3</sup> and this image has been made on a range of materials from each of the major film suppliers. Kodak supplies the image on both Kodachrome and Ektachrome transparency material, and Fuji and Agfa supply it on Fujichrome and Agfachrome respectively. Each of the vendors also supplies the same image on their print materials. (Konica is also heavily involved in the development of this target but currently has no plan to market it.)

However, those expensive products result in low usability for end users, "calibrated" targets especially. The reasons are as follows. First, ISO 12641 is strict with the chroma values, which are not easily achieved by other manufacturers beyond above four vendors. Significantly, the colorimetric aim values of the target are designed around the characteristics of color transparency film and photographic paper. For example, manufacturers shall test the maximum chroma at each defined lightness and hue, if they use different photographic materials. Second, the traditional manufacturing approach is very time-consuming during making the intermediate film. Third, the techniques for process control in producing targets are expertise and the requirements for achieving satisfactory control are rigorous. Final, it is originally intended that vendor would

---

Correspondence: Chao-hua Wen; OES/ITRI, S010, B51, 195-8, Sec. 4, Chung-Hsing Rd., Chu-tung, Hsin-chu 310, Taiwan, R.O.C.; Email: h880021@itri.org.tw; Telephone: 886-3-5913714; Fax: 886-3-5829781.



provide the "calibrated" target so that the user could know the tristimulus values of each patch on that specific target. This would be necessary for those who require device independent calibrations and have no appropriate color measuring facilities themselves. However, early experiences suggested that, for some vendors at least, the sample variation within a batch was sufficiently small that use of a batch calibration data set was adequate for all but the most critical users. For this reason, at least one of the vendors (Kodak) has chosen to make batch average data for their targets available from an Internet site: <ftp://ftp.kodak.com/gastds/Q60DATA/>.

To overcome these barriers, this study proposes a new approach of calibration target design and manufacture. First, we discuss with the requirements of the calibration target. Second, this article describes how to design the calibration target that differs from Maier and Rinehart<sup>3</sup> and specifies how to modify the aim values of the ISO/FDIS 12641 color patches for meeting process capability of color reproduction. Third, we draw the processes of conversion of numerical CIELCH data for each patch into a sRGB-mode electronic image file. Fourth, this study explains how to use dynamic subgroup linear interpolation (DLSI) and statistic process control (SPC) to achieve high process quality of printing calibration targets and cost-down. Finally, a brief of discussion and conclusions has been made for future researches.

## 2. REQUIREMENTS OF THE CALIBRATION TARGET

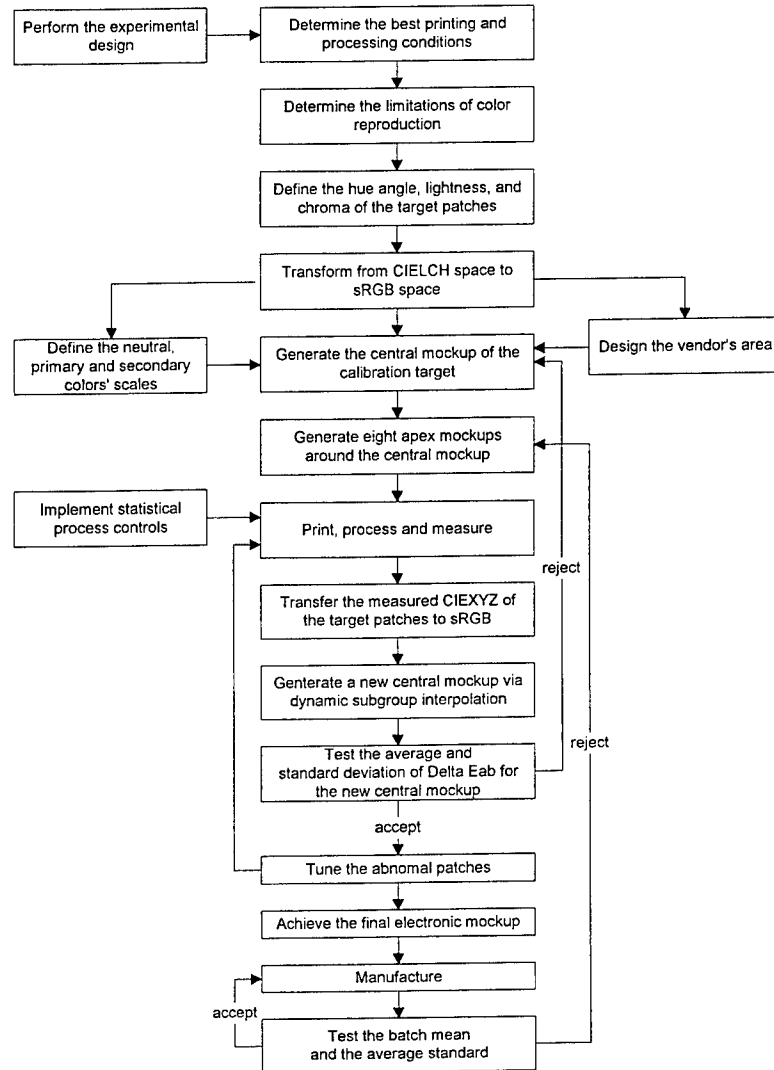
This research primarily addresses itself on the reflection target. IT8.7/2-1993 consists of a series of single dye scales (cyan, magenta and yellow) with equivalent two- and three-dye combinations.<sup>4</sup> These scales are particularly useful in setting up many scanners since the resultant colors are similar to the regions of color space in which the scanners permit independent. The remaining colors consist of 12 samples at each of 12 hues. The samples include four chroma levels at each of three lightness levels. Each of the lightness intervals and each of chroma intervals are approximately equally spaced, and each of maximum chroma which material can be achieved is produced at each lightness and hue level. A 22 step gray scale is also included. Apart from above specific values, the target also contains the  $D_{min}$  and  $D_{max}$  and 36 vendor selected colors. Therefore, IT8.7/2-1993 provides altogether 288 colors.

Thus, the calibrated color target provides known tristimulus values which cover, more or less uniformly, the full colour gamut of the specific material on which the target has been imaged, as an input to the color scanner characterization. In one word, the target must have a set of colored patches that cover the color gamut of the photographic material and important and common hues. The colored patches could also provide a guide for making the color correction adjustments on the electronic color separation scanner and an orderly pattern to facilitate automated evaluation of predetermined values stored in a look-up table. Kang also concluded that the position of the color used for training is more important in the color interpolation within a given gamut rather than the number of colors.<sup>5</sup>

## 3. DESIGN AND SPECIFICATION OF THE CALIBRATION TARGET

Figure 1 demonstrates the architecture of calibration target design and production, as follows. First, the best printing and processing conditions will be determined via the experimental design. Second, color gamuts will be illustrated for understanding the limitation of color reproduction by a field study, not by computing methods<sup>6,7</sup>. Third, the color values of target have been specified that according to experimental color gamuts. Fourth, color transform from CIELCH to sRGB will be described. Fifth, the relevant target mockups will be created. Sixth, the quality will be controlled by SPC during daily setup. Seventh, the method of DLSI will be adopted to refine the electronic target mockup. Final, an experiment will show how to supervise the quality of photographic printer and processor for meeting the aim values. This chapter focuses on preparatory works, color theories, calibration target design and specification of the targets from the first stage to the fifth stage. The remainders will be discussed on the next chapter.

This study applies the CIE 1976 ( $L^*a^*b^*$ ) color space or the CIELAB color space to design the color calibration target.<sup>8</sup> Uniform spacing in hue, lightness and chroma, and tolerance in terms of differences in these parameters ( $\Delta E_{ab}$ ) is believed to provide a reasonable distribution of target patches in the most effective manner. In this color space,  $L^*$  represents lightness and  $a^*-b^*$  coordinates indicate the hue and chroma information. Hewlett-Packard and Microsoft proposed the addition of support for a standard color space, sRGB, within the Microsoft operating systems, HP products, the Internet, and all other interested vendors. The aim of this color space is to complement the current color management strategies, particularly ICC, by enabling a simple method of handling color in operating systems and the WWW.<sup>9</sup> This method utilizes a simple and robust color space definition that will be applied.



**Figure 1.** The architecture of calibration target design and production

### 3.1. Determination of printing and processing conditions

Today's photographic market is rapidly changing. This is primarily due to internal developments and growing influence of external factors. These factors are attribute to digital electronics, communications, entertainment and leisure-time opportunities. Efforts to render digital images onto silver-halide paper and film for commercial are quite recent.

Recently, Noritsu's QSS23-HRCRT with the high resolution CRT printer has become available in the market. HRCRT is short for Hyper Resolution Cathode Ray Tube. This is a digital printing engine with higher quality of 500 or 300 dpi prints in minilab industry and suitable for high-end users. We use 500 dpi for high performance of the HRCRT in this study. The HRCRT electron beam lights up three phosphor lines: blue, green, and red. The paper is then pressed against the HRCRT tube faceplate (without any lens) and exposed. The processing capacity of HRCRT printer is 122 prints per hour for size of 127mm (width)  $\times$  178mm (advance length) under 500dpi.

We use Kodak Professional Digital Paper Type 2976 that is designed for digital printers rather than Royal series paper that is designed for optical exposure. Moreover, we use Kodak Ektacolor RA chemical for Process RA-4 to process the digital paper under normal replenishment rate. In addition, gamma of the HRCRT plug-in driver has been tested and settled on 1.8 by way of printing gray scale at varied gamma levels.

### 3.2. Color gamut of color reproduction

Figure 2 represents the color gamuts of Kodak digital paper type 2976, Kodak Royal VII paper and sRGB. This plot shows all points on the  $a^*-b^*$  plane without information about the  $L^*$ . The sRGB's color gamut is obtained from calculating all points of RGB  $32 \times 32 \times 32$ , and color gamuts of two kinds of paper are measured from printing all points of RGB  $17 \times 17 \times 17$  respectively. The orange enclosure is the gamut of Kodak professional digital paper type 2976. The black enclosure is the gamut of Kodak Royal 7 Color paper. And the blue one is the gamut of sRGB. Figure 2 also shows that the gamut of sRGB is wider than papers.

One phenomenon is worthy to be mentioned that papers give larger gamuts in yellow and cyan colors than sRGB. Unfortunately, Figure 2 reveals accurate hue information, but without the lightness information. The plot of  $L^*-C^*$  provides a suitable second graph, as show in Figure 3, to represent CIELAB color space. Although this plot contains only colors at the specified hue angle, this diagram is most useful for depicting the color gamut of a system at a particular hue angle. Figure 3 also demonstrates that some specified colors in IT8.7/2-1993 are outside of color gamut of digital paper after printing and processing, such as F3, G3, H3, and L3. With these color gamuts, this research can now define color patches for the calibration target and that will be discussed in the next section.

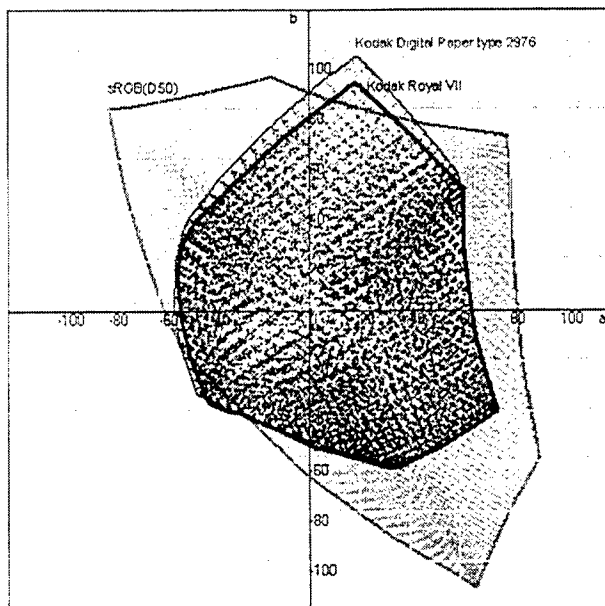


Figure 2. Comparisons of color gamuts among Kodak digital paper type 2976, Kodak Royal VII paper and sRGB

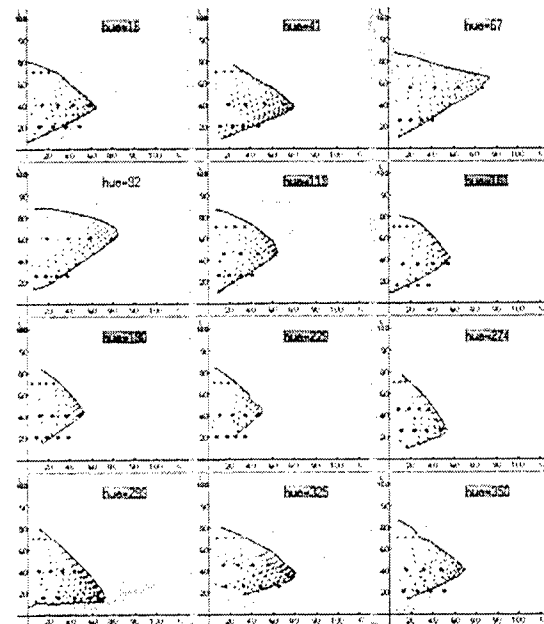


Figure 3.  $L^*-C^*$  Diagrams showing the gamuts of digital paper (dark) and sRGB (light) for 12 hue angles (the block points are aim values specified in IT8.7/2-1993)

### 3.3. Specification of target patches

Good color photography is expected to achieve large color gamuts and stable neutral grays. However there is a general tendency that three dyes giving large color gamuts give unstable neutral grays. Modern color photography tends to adopt the three dyes giving large color gamuts at the expense of the stable neutral grays. Therefore, the first thing is to define the neutral scale. Anyway, we follow the definition of neutral scale in IT8.7/2-1993. The neutral scale has equal visual intervals, which in CIELAB color space means equal interval in  $L^*$ . The range in densities is from white to black. For this scale to be neutral,  $a^* = 0$  and  $b^* = 0$ . This study also complies with the specifications of the primary color, the secondary color and black scales in ISO 12641.

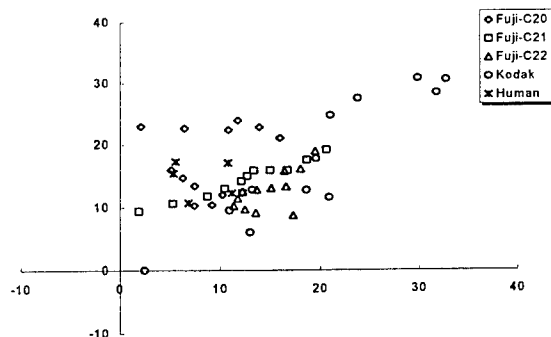
In accordance with results of section 3.2, this paper modifies parts of the definition of IT8.7/2-1993 about color gamut mapping. We retained the hue information as well as IT8.7/2-1993. This research also kept the  $L^*$  value that is defined in IT8.7/2-1993 as far as possible. Additionally, this study performs in conformity with the essence of chroma specification in IT8.7/2-1993, which is equal space in chroma. All color gamut-mapping patches are listed in Table 1.

A "vendor-optional" area is provided so that different target manufacturer can add unique elements by self-determination. Here, we investigate the facial colors of Kodak Q60 color input target, Fuji color target and real humankind in CIELAB color space, as shown in Figure 4. Diamonds represent the 20 column patches of Fuji color target, squares denote the 21 column patches of Fuji color target, triangles show the 22 column patches of Fuji color target. Circles mean 12 patches from I20 to L22 of Kodak Q60 color input target. And asterisks indicate skin colors (forehead/cheek) of African, Arabian, Caucasian, Japanese and Vietnamese respectively. Fortunately, sorting those colors by red after transforming facial colors from CIELAB to sRGB (discussed in next section) that clue us in facial color self-determination, depicted in Figure 5. Diamond denotes red, square denotes green and triangle denotes blue. Two arrows point out the two outliers in blue channel. The reason of sorting by red channel is that people's blood is hot red. And Figure 5 evidences this hypothesis. Red is rapidly increased between 0 and 150 and is convergent between 200 and 220. Generally, blue have the same trend as green and blue is smaller than green. Next, we used polynomial curve fitting to find the coefficients of three appropriate polynomials that fit the R, G and B data individually. Finally, we calculate the values of the polynomials evaluated at equal intervals that are equal divided the number of effective sampling facial colors into the number of demanded patches.

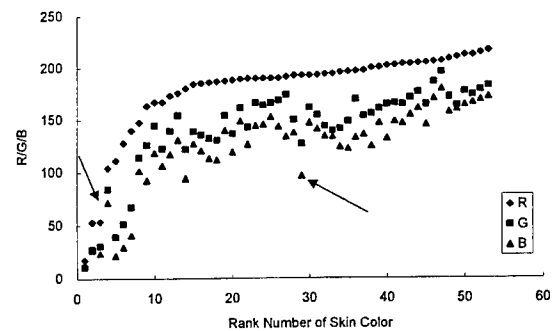
**Table 1.** Hue angle, lightness and chroma

Row	Hue Angle	L1					L2					L3				
			C1	C2	C3	C4		C1	C2	C3	C4		C1	C2	C3	C4
A	16	20	9*	18*	27*	36*	40	15	30	44	62*	70	7	14	21	28*
B	41	20	10*	20*	30*	40*	40	20	36	54	76*	70	8	16	24	32*
C	67	25	9*	18*	27*	36*	55	22	44	66	81*	75	10	20	30	40*
D	92	25	7*	14*	21*	28*	60	20	40	60	80*	80	10	21	31	41*
E	119	25	8*	16*	24*	33*	45	14*	28*	42*	56*	70	9	18	27	36*
F	161	20*	6*	12*	18*	25*	35	11*	22*	33*	43*	70	6	12	18	24*
G	190	20	6*	12*	18*	23*	40	13	25	38	49*	70	6	13	19	26*
H	229	20	6*	12*	18*	23*	40	12	24	36	45*	70	7	13	20	27*
I	274	25	12*	24*	36*	48*	45	9	19	28	35*	70	5	10	15	20*
J	299	15	14*	28*	42*	56*	40	11	22	33	51*	70	6	11	17	22*
K	325	25	14*	28*	42*	56*	45	14	28	42	70*	70	8	16	24	32*
L	350	20	9*	18*	27*	35*	40	16	32	48	66*	70	8	15	22	31*

Asterisks denote modified or self-determined color patch values



**Figure 4.** Collected facial colors in CIELAB color space.



**Figure 5.** The appearance of facial colors, which are sorted by red after transforming CIELAB to sRGB

### 3.4. Generation of the electronic target mockup

The electronic target mockup is for digital printer. Because the QSS2301-HRCRT only receives the image file of 24-bit RGB color mode, transforming color spaces from CIELCH to sRGB at the D50 white point should be discussed firstly. In general, many publishers only write the calculation processes of color space transformation from XYZ to CIELCH. Since Stokes et al. proposed the sRGB color space,<sup>9</sup> the sRGB has been widely introduced to the color management system. However, the D50 white point of the profile connection space is included in the header of the profile, but the white point is set to D65 of the sRGB monitor. Recently, Nielsen and Stokes proposed a 3x3 matrix transforming sRGB to XYZ at D50, which is the product of the reduced Bradford chromatic adaptation matrix and the sRGB matrix at D65, to clear up the confusion.<sup>10</sup>

This study developed a generator of the electronic semi-mockup for the calibration target. The generator had two main functions. One is that transforms the 288 patch color values from XYZ to sRGB, and makes up an initial/central semi-mockup which size is 416x288 pixels or produces an apex semi-mockup which is based on both of the setting interval and the central semi-mockup. The intention of apex semi-mockups prepared for DSLI will be described in next chapter. The other is that performs DSLI. Figure 6 shows the initial electronic semi-mockup image. This is not a finish mockup, but it is still necessary to edit the image file for meeting the specifications of QSS2301-HRCRT, such as image resolution, image size, labels of the calibration target, and so on. We used Photoshop 5.0 as a tool for editing the image file of semi-mockup to finish the electronic target mockup.

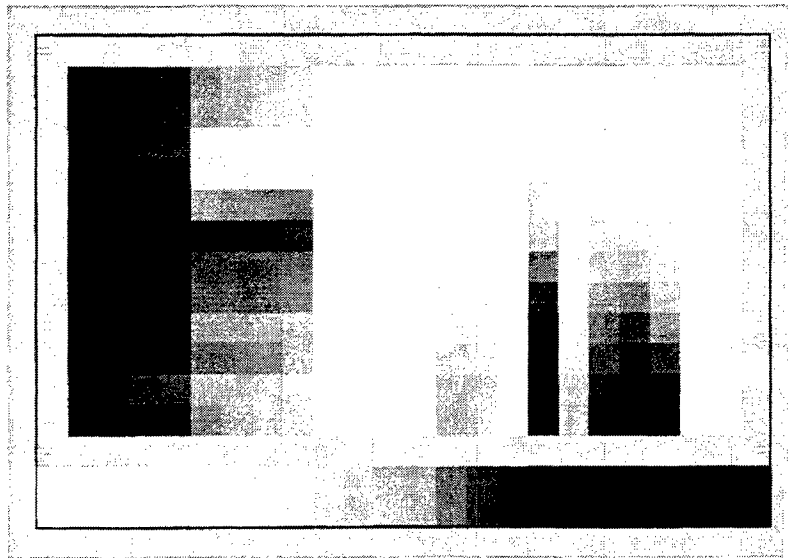


Figure 6. The initial electronic semi-mockup

## 4. QUALITY OF THE PRODUCTION PROCESS

Quality control in manufacturing the calibration target is a very important issue. ISO 12641 specifies that the "un-calibrated" target patches contained with A1 through L3, A5 through L7, and A9 through L11, 99% shall be within 10 delta Eab\* of the aim values for all targets manufactured. ISO 12641 also specifies that 99% of the "un-calibrated" target patches within the manufacturing batch shall be within 5 delta Eab\* of the reference for each manufacturing batch. The reference are reported batch means for patches A1 through L19, Dmin and Dmax, and the aim values specified in this standard for 22-step neutral scale.<sup>2</sup> However, those manufacturing tolerances have been set which are capable of being achieved over a significant number of targets and have been considered how to meet the objective of minimizing variations as far as reasonable. Following sections will discuss how to implement DSLI to refine a "gold" electronic target mockup and how to introduce SPC into supervising the printing and processing processes. Finally, experimental results demonstrated one general principle for controlling the calibration target manufacturing process.

#### 4.1. Dynamic subgroup linear interpolation

Color space transitions are frequently used in image processing as mention in the previous chapter. A image encoded in a device-independent representation, such as CIELAB, CIELUV and XYZ, needs to be transformed to the device space of the monitor before it can be displayed and to the colorants of printer before it can be printed. Unfortunately, the interaction of light with the dyes and pigments of practical printers to form colors is more complex than the color-forming mechanisms of CRTs, making it more difficult to construct an accurate mathematical model for a printer.<sup>11</sup> It is possible to specify arbitrary transfer functions in terms of three-dimensional lookup tables, but the requirement of huge storage and the unpredictable quality of printing are shortcomings of this approach. Here, we propose DSLI to overcome above drawbacks.

The goal of this research is to produce accuracy 288 patches of the calibration target, not to improve image quality. After color space transforming from CIELCH to sRGB, we can acquire the results that are near but are not exact. Thus, those principles inspired the design concept of dynamic subgroup linear interpolation. The approach is not necessary to construct an accurate mathematical model for QSS23-HRCRT printer or build the huge-volume lookup table for high printing quality. But the approach creates one subgroup for each patch. The subgroup is called body-centered-cubic packing that consists of nine patches, one center and eight apices. This is shown in Figure 7. Intuitively, the initial electronic mockup consists of 288 patches, one center and eight apices. This is shown in Figure 7. Intuitively, the initial electronic mockup consists of 288 patches, one center and eight apices. The size of the subgroup depends on the magnitude of color difference between the central mockup and its aim values. Therefore, this paper simply stores the color values of 288 subspaces for the output function evaluated at discrete points in the input space.

Following printing and processing those nine electronic mockups with QSS2301-HRCRT, we measured them by Gretag Macbeth Spectrolino & SpectroScan. The color measurements were performed on a black backing in accordance with ISO 5/4 under the conditions 45/0 geometry, no polarization, no filter, 2° observer, D50, XYZ system. Next we transforms the measured XYZ values of nine mockups to sRGB space. The simple linear interpolation in one dimension is enough to evaluate the function between inputs and outputs in this study. The procedures are as follows: (1) calculate the individual RGB color difference between output and its aim value for each subgroup, (2) search the minimum of RGB difference and determine the direction of interpolation, and (3) evaluate the new value. Because the complex interaction of paper dyes, the algorithm should be noted when small value occurs.

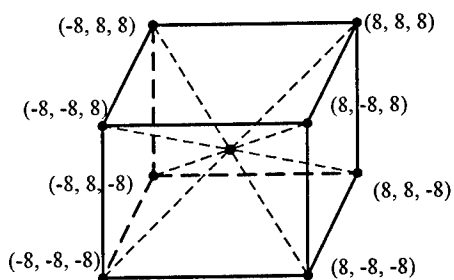


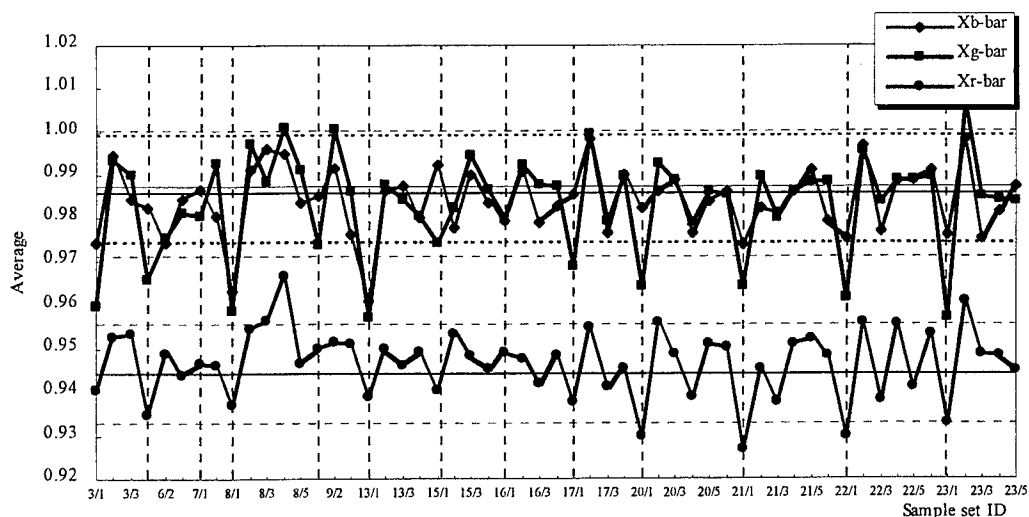
Figure 7. One example of body-centered-cubic packing

#### 4.2. Supervision of photographic printer and processor

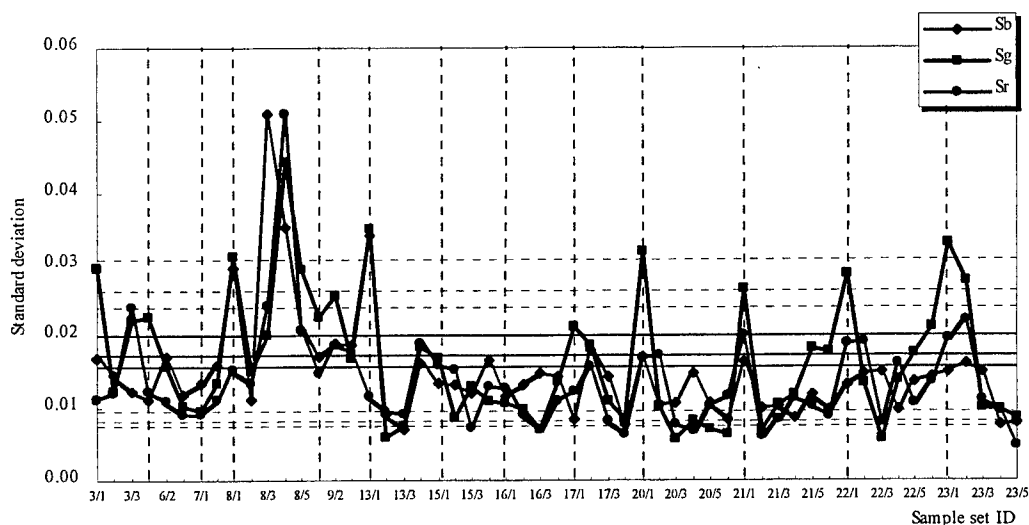
From a photographic profession's viewpoint, color balance is a fundamental requirement for getting the best quality of color reproduction. Observation the reproduction of the gray scale is frequently used in color balance. This study adopted the 18 step daily setup to control the exposure of QSS23-HRCRT. Although the function of daily setup is automatic that we cannot intervene, the process stability of daily setup is a critical factor that may supervise the quality of chemicals and archive the high output quality. We develop a method from the concept of SPC into monitoring the primary quality during research and design phase. However, we believe that the method is adequate to the manufacturing lineup.

During daily setup, Densities of 18 steps are measured by the inner densitometer of QSS23-HRCRT and are displayed in RGB. These raw data are miscellaneous. The most commonly used measure of location or tendency is the mean and two measures of dispersion extremely used are the range and the standard deviation.<sup>12</sup> For simplifying and characterizing, the measures are not classical statistic forms any more, but they still contain both basic and specific information. In this study, for each sample of 18 step density, the average and standard deviation (the reference is the overall mean density of each step, not the average) were computed for each channel. Two control charts had been drawn to monitor the materials, manufacturing processes, and so forth. Figure 8 and Figure 9 illustrate an X-bar chart for monitoring the location parameter

and an S (standard deviation) chart for monitoring the dispersion parameter respectively. From the past experience, one time of daily setup is not enough to steady the printing and processing of QSS23-HRCRT. This results in a varied of sampling sizes for one test. In general, performing three to eight times daily setup are necessary for good quality, shown in Figure 9. The first time of daily setup has the largest standard deviation. However, continual daily setups could help decrease the standard deviation. Finally, results revealed that we should control the standard deviation below its average as far as possible.



**Figure 8.** Control chart for averages, X-bars, with 3-sigma control limits for 13 running days



**Figure 9.** Control chart for standard deviations, S's, with 3-sigma control limits for 13 running days

## 5. RESULTS AND DISCUSSIONS

### 5.1. Verification of DSLI

Figure 10 summarizes the overall  $\Delta E_{ab}^*$  of each patch of two central mockups in the initial stage and the third DSLI stage. During performing DSLI three times, the interval settings of DSLI, that are the size of subgroup, were 8, 4 and 2, respectively. Results showed that DSLI could improve the color accuracy of all patches. However, there were some patches did not act as well as the overwhelming majority of patches. For example, the values of  $\Delta E_{ab}^*$  of C7, G7, G8, H7, K4 and L4 patch were over 6. One interesting phenomenon was that all these abnormal patches were close to high chroma. In other words, the usability of DSLI was limited to these color patches in this study. This phenomenon will be investigated in the following solution. DSLI did not consider that chroma information at the same level of hue and lightness. For abnormal patches, we adopted other accuracy patches, which were at the same level of hue and lightness in the final stage. For instance, we use polynomial curve fitting function to fit C5, C6 and C8 in RGB mode and evaluate the polynomial at C7. In general, only one color channel needs processing.

Figure 11 shows averages and maximums of  $\Delta E_{ab}^*$  resulting from the initial stage to the final stage. The average of  $\Delta E_{ab}^*$  was 9.01 and the standard deviation of  $\Delta E_{ab}^*$  was 5.51 in the initial stage, which did not employ DSLI. Through using DSLI once, the average of  $\Delta E_{ab}^*$  was 3.73 and the standard deviation of  $\Delta E_{ab}^*$  was 3.21, and so on. After performing DSLI three times, we called the third DSLI stage, the average of  $\Delta E_{ab}^*$  was downward to 1.69 and the standard deviation of  $\Delta E_{ab}^*$  was also downward to 1.46. These results are listed in Table 2.

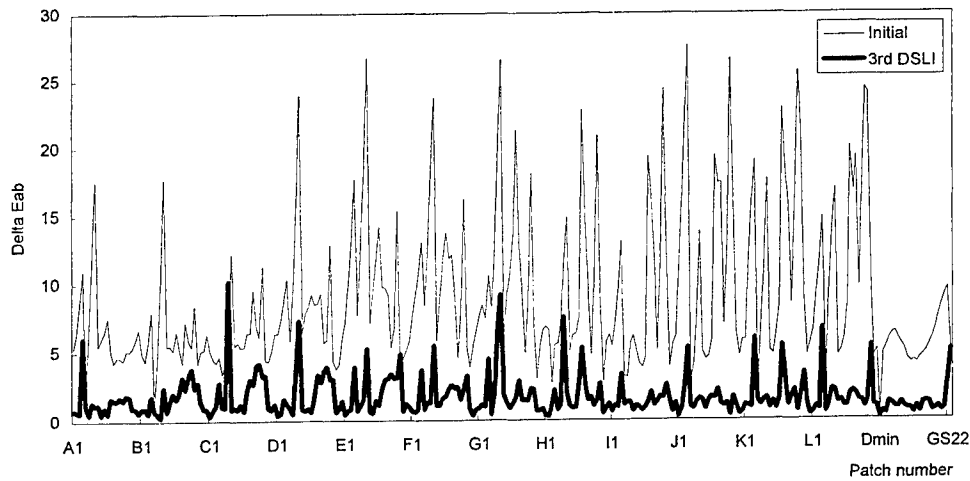
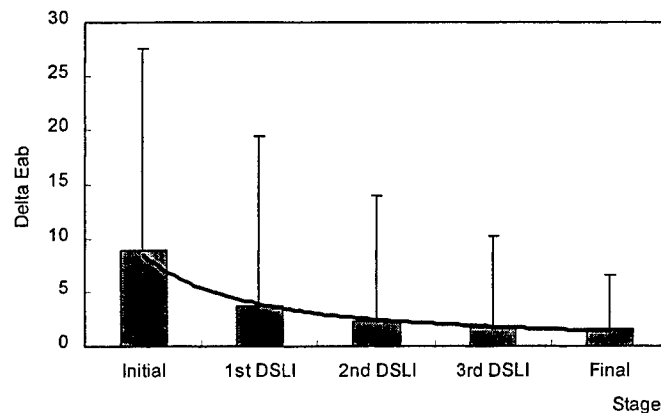


Figure 10. Comparisons of the distributions of  $\Delta E_{ab}^*$  of the central mockup between the initial and third stages

Table 2. Statistic analysis results form the initial stage to the final stage

$\Delta E_{ab}^*$	Initial stage	1st DSLI stage	2nd DSLI stage	3rd DSLI stage	Final non-DSL stage
Maximum	27.58	19.53	14.00	10.18	6.66
Minimum	0.33	0.22	0.24	0.16	0.05
Average	9.01	3.73	2.35	1.69	1.554
Standard deviation	5.51	3.21	1.73	1.46	1.31

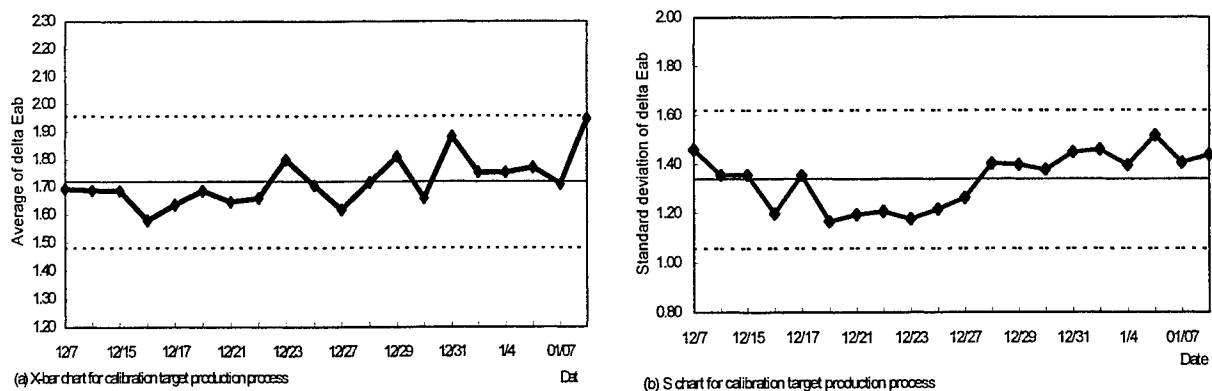




**Figure 11.** Efficiency of implementing DSLI in decreasing the average and the maximum of delta Eab\*

## 5.2. Long-term testing

For long-term testing the production process, in this experiment, we randomly sampled one target and measured it from 20 produced calibration targets per workday. The measurement conditions were likely to 4.1. The average and the standard deviation of delta Eab\* for one sample were computed. Averages and standard deviations of delta Eab\* from 20 samples are collected, as shown in Figure 12. Results demonstrated that we could control averages of delta Eab\* between 1.48 and 1.96 and standard deviations between 1.62 and 1.06 (Figure 12 (a)). The overall average of delta Eab\* was 1.72 and the average of standard deviations of delta Eab\* was 1.34 (Figure 12 (b)). A value of 1.34 means that 99% of the calibration targets will be with a delta Eab\* of 4.489 (3.35× average of standard deviation) of the manufacturing aim values.<sup>13</sup> During the long-term testing, the fluctuations of average delta Eab\* became larger day by day. As Figure 12 (b) depicts, standard deviations were over the average of over all standard deviations in a later period. Theoretically, if manufacturing process happens to be out of control, DSLI is still practical.



**Figure 12.** X-bar and S chart for calibration target production process

## 6. CONCLUSION

This paper has discussed how to design and produce the calibration targets for digital input color devices. In addition, this study has thoroughly examined the effectiveness of DSLI and has capably controlled the quality of calibration targets from a field experiment. We infer that the approach is also suitable for manufacturing transparent targets. Up to present, although

some researches have focused on development of virtual spectral color target, there are still some limitations, such as metamerism.<sup>13</sup> Real color patches, made of photographic materials, have smoother spectral shapes than the virtual color patches. However, because this study has specified some self-determinations differ with ISO 12641, the feasibility of the calibration target will be compared with other vendors in future works.

## REFERENCE

1. Tony Johnson, "Methods for characterizing colour scanners and digital cameras," *Displays*, **6(4)**, pp.183-191 (1996).
2. ISO, *ISO/FDIS 12641 Graphic Technology—Prepress Digital Data Exchange—Colour Targets for Input Scanner Calibration*, 1997.
3. T. G. Maier and C. E. Rinehart, "Design Criteria for an Input Color Scanner Evaluation Test Object," *TAGA Proceedings*, pp. 469-483, 1988.
4. ANSI, *ANSI IT8.7/2-1993 Graphic technology—Color reflection target for input scanner calibration*.
5. H. R. Kang, "Color scanner calibration of reflected samples," *SPIE Vol. 1670 Color Hard Copy and Graphic Arts*, pp. 468-477, 1992.
6. Ohta, N., "The Color Gamut Obtainable by the Combination of Subtractive Color Dyes. V. Optimum Absorption Bands as Defined by Nonlinear Optimization Technique." *Journal of Imaging Science*, **30(1)**, pp. 9-12, 1986.
7. Inui, M., "A Fast Algorithm for Computing the Colour Gamut of Subtractive Colour Mixture." *Journal of Photographic Science*, **38(4,5)**, pp. 163-164, 1990.
8. CIE 15.2, *Colorimetry*, Second edition, 1986.
9. M. Stokes, M. Anderson, S. Chandrasekar and R. Motta, "A standard default color space for the Internet – sRGB," Version 1.10, <http://www.w3.org/Graphics/Color/SRGB.html>, 1996.
10. M. Nielsen and M. Stokes, "The Creation of the sRGB ICC profile," *Proceedings of IS&T/SID 1999 Color Imaging Conference: Color, Systems, and Applications*, pp. 253-257, 1999.
11. J. M. Kasson, "Performing color space conversions with three-dimension linear interpolation," *Journal of Electronic Imaging*, **4(3)**, pp. 226-250, 1995.
12. W. J. Kolarik, *Creating Quality: Concepts, Systems, Strategies, and Tools*, McGraw-Hill, New York, 1995.
13. F. K. Dolezalek, "Appraisal of production run fluctuations from color measurements in the image," *TAGA Proceedings*, pp. 154-164, 1994.
14. H. Kotera, H. S. Chen and R. Saito, "Generation of virtual spectral target and its applications," *Proceedings of The Seventh IS&T/SID Color Image Conference: Color Science, Systems and Applications*, pp. 36-41, November 16-19, Scottsdale, Arizona, USA, 1999.

# An implementation of scanner ICC profile generator

Yi-Ching Liaw and Chun-Yen Chen

Industrial Technology Research Institute, Opto-Electronics & Systems Laboratory,  
Hsinchu, Taiwan, R.O.C.

## ABSTRACT

As the popularity of color peripheral devices grows, the problem of color inconsistency among color devices becomes more and more important. The ICC Profile, specified by the International Color Consortium (ICC), is a reasonable solution to achieve color consistency. This paper describes an implementation of automatic scanner ICC Profile generator. The generator mainly consists of a reference-file-parser, a target-image-analyzer, and a profile-parameters-evaluator. The reference-file-parser retrieves the XYZ and Lab values of color patches from the ANSI CGATS.5 compliance file. The RGB values of color patches are automatically extracted from the scanned ANSI IT8.7/2 target by the target-image-analyzer. Finally, the parameters that are required for profile construction are generated using the profile-parameters-evaluator. Based on the test of 4 types of scanners, the average  $\Delta E^*_{ab}$  of less than 3 is achieved..

**Keywords:** Scanner ICC Profile, Color Calibration

## 1. INTRODUCTION

As the popularity of color peripheral devices grows, the problem of color inconsistency among color devices becomes more and more important. The ICC Profile [1][2], specified by the International Color Consortium (ICC), is a reasonable solution to achieve color consistency. The major purpose of ICC Profiles is to characterize the color devices and to provide the color space transformation information between color devices and CIE standard. The CIE XYZ and CIE Lab are selected by the ICC as the intermediate color spaces and are called the Profile Connection Space (PCS). The relation between PCS and ICC Profiles is shown in figure 1.

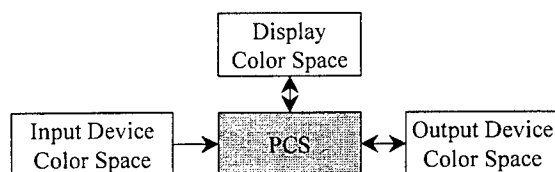


Figure 1. The relation between PCS and ICC profiles

There are two color space transfer models are defined for the device ICC profiles. One is called three-component matrix-based model and another is called N-component Look-Up-Table (LUT)-based model. In general, each model suits for different kinds of devices. For the scanner ICC profile, both the two models can be used and are implemented in our scanner ICC profile generator. The two models are described as follows.

**Three-component matrix-based model:** The three-component matrix-based model mainly consists of a red Tone-Reproduction-Curve (TRC), a green TRC, a blue TRC, a red colorant, a green colorant, and a blue colorant. Where the RGB TRCs are used to linearize the tone curves of RGB and the RGB colorants are used to translate the linearized RGB values to XYZ values. For the given RGB values retrieved from the scanner, the RGB values can be translated to XYZ by using the following equations.

$$linear_r = TRC_r[Device_r] \quad (1)$$

$$linear_g = TRC_g[Device_g] \quad (2)$$

$$linear_b = TRC_b[Device_b] \quad (3)$$

$$\begin{bmatrix} PCS_x \\ PCS_y \\ PCS_z \end{bmatrix} = \begin{bmatrix} redColorant_x & greenColorant_x & blueColorant_x \\ redColorant_y & greenColorant_y & blueColorant_y \\ redColorant_z & greenColorant_z & blueColorant_z \end{bmatrix} \begin{bmatrix} linear_r \\ linear_g \\ linear_b \end{bmatrix} \quad (4)$$

Where the  $Device_r$ ,  $Device_g$ ,  $Device_b$  are the RGB values got from the scanner, the  $linear_r$ ,  $linear_g$ ,  $linear_b$  are the linearized RGB values, and the  $PCS_x$ ,  $PCS_y$ ,  $PCS_z$  are the corresponding XYZ values of the RGB values got from the scanner. To use the three-component matrix-based model for color space transformation, the RGB values of scanner are firstly mapped to the linear version using the individual TRC. After that, the relative XYZ values of the linearized RGB values can be calculated by multiplying the RGB colorant with the linearized RGB values.

**N-component LUT-based model:** The N-component LUT-based model used in the scanner ICC profile consists of RGB TRCs, a three-dimensional LUT, and three linearization curves. The RGB TRCs are used to linearize the input channels. The three-dimensional LUT is used to translate the color space from scanner color space to CIE Lab. The three linearization curves are used to linearize the output channels. The N-component LUT-based model for scanner ICC profile is shown in the figure 2.

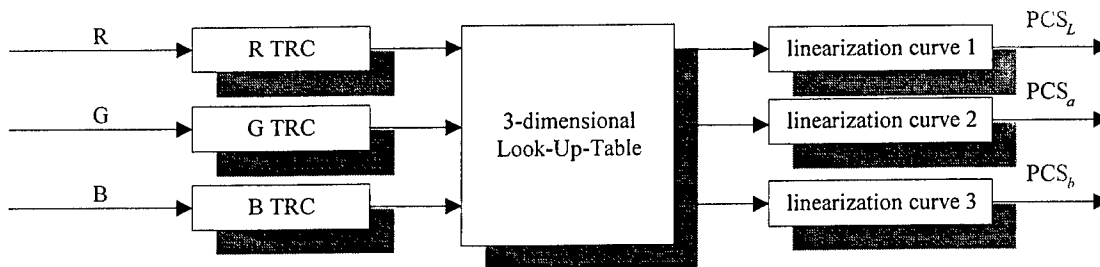


Figure 2. The N-component LUT-based model for scanner ICC profile

In this paper, two models described above are implemented in the proposed scanner ICC profile generator. Except that, the generation process of scanner ICC profile is also be automated by introducing the reference-file-parser and the target-image-analyzer. The rest of this paper is organized as follows. The system architecture of scanner ICC profile generator is given in section 2. The implementation details of reference file parser, target image analyzer, and profile parameter evaluator are described in section 3. The experimental results are listed in section 4. Finally, the conclusion are given in section 5.

## 2. SYSTEM ARCHITECTURE OF THE SCANNER ICC PROFILE GENERATOR

To generate the scanner ICC profile, the differences between the scanner color space and CIE standard must be discovered. The IT8.7/1 [3] and IT8.7/2 [4] targets are the standard color targets defined by ANSI and contains variety color patches that can be used to analyze the color characteristics of scanner. Figure 3 shows the layout of IT8.7/1 and IT8.7/2. In the bottom of the target, a gray-bar with 24 patches is given to characterize the tone representation of input devices. The 12 rows by 22 columns color patches are given in the top of the target can be use to characterize the color features of input devices.

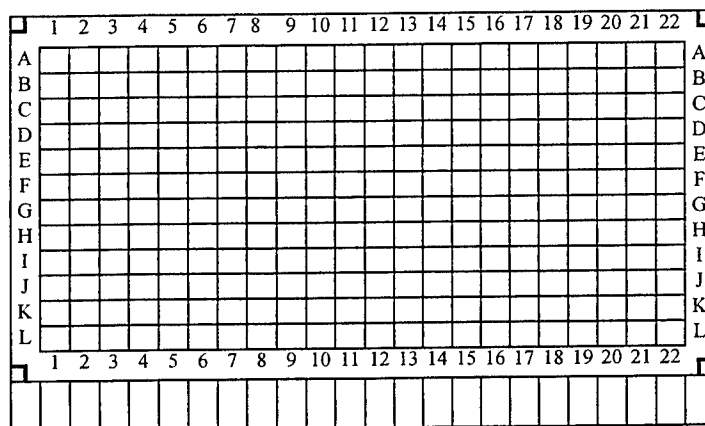


Figure 3. The layout of IT8.7/1 or IT8.7/2

By using the IT8.7/1 or IT8.7/2 target for generating the scanner ICC profile, the RGB values of color patches from scanned target and the XYZ values of color patches from colorimetric are required. Once the RGB and XYZ values of the color patches are acquired, the parameters required for constructing the scanner ICC profile can be evaluated. The system architecture of the scanner ICC Profile generator is shown in the following.

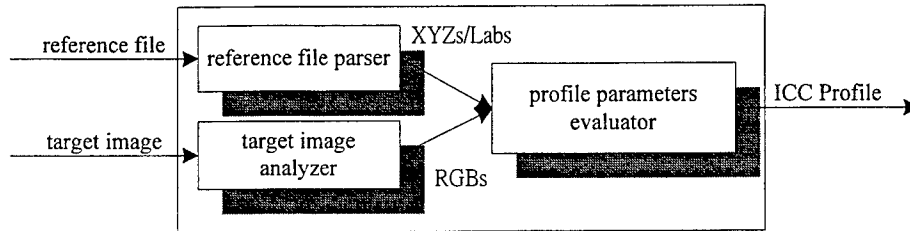


Figure 4. The system architecture of scanner ICC profile generator

Where the reference file is a ANSI CGATS.5 [5] compliance file for storing the measured XYZ and Lab values of color patches of the IT8.7/1 or IT8.7/2 target and the target image is an image of the IT8.7/1 or IT8.7/2 target scanned by the testing scanner. The generator mainly consists of a reference-file-parser, a target-image-analyzer, and a profile-parameters-evaluator. The reference-file-parser is used to retrieve the XYZ and Lab values of color patches from reference file. The target-image-analyzer is used to extract the RGB values of color patches from the target image. Once the XYZ, Lab, and RGB values of color patches are acquired, the parameters required for three-component matrix-based model and the N-component LUT-based model can be evaluated by the profile-parameters-evaluator.

### 3. IMPLEMENTATION OF THE SCANNER ICC PROFILE GENERATOR

In this section, we will present the implementation of scanner ICC profile generator more detail. The reference-file-parser, target-image-analyzer, and profile-parameters-evaluator are described in the following.

#### 3.1. reference-file-parser

The reference-file-parser is used to retrieve the XYZ and Lab values of color patches from the ANSI CGATS5 compliance file. The ANSI CGATS5 standard specifies a methodology for taking spectral measurements and making colorimetric computations as well as defines a data exchange format for storing the measured color data. Figure 5 shows an example of ANSI CGATS5 compliance file.

IT8.7/2						
ORIGINATOR	"S010/OES/ITRI"					
DESCRIPTOR	"Color Target (Illumination=D50 ObserverAngle=2)"					
CREATED	"4/16/2000"					
NUMBER_OF_FIELDS	7					
BEGIN_DATA_FORMAT						
SampleName	XYZ_X	XYZ_Y	XYZ_Z	LAB_L	LAB_A	LAB_B
END_DATA_FORMAT						
NUMBER_OF_SETS	288					
BEGIN_DATA						
A1	4.40	3.76	2.65	22.88	11.15	3.44
A2	5.15	3.58	2.14	22.25	23.51	6.75
	...					
Dmax	0.47	0.57	0.47	5.14	-3.32	-0.06
END_DATA						

Figure 5. An example of ANSI CGATS5 compliance file

The algorithm for retrieving the reference file is presented as follows.

### *The algorithm of reference-file-analyzer*

- (1) File validity checking: The first 7 characters in the file must be the "IT8.7/1" or "IT8.7/2".
- (2) Data format retrieving: The arrangement information of XYZ and Lab values are encapsulated by the "BEGIN\_DATA\_FORMAT" and "END\_DATA\_FORMAT" keywords. Before retrieving the measured data of target, the arrangement information of XYZ and Lab values must be known.
- (3) Measured data retrieving: The XYZ and Lab values encapsulated by the "BEGIN\_DATA" and "END\_DATA" keywords can be retrieved using the arrangement information of the XYZ and Lab values.
- (4) Data validity checking: To ensure the retrieved XYZ and Lab values are the correct XYZ and Lab values of IT8.7/1 or IT8.7/2 target. The XYZ and Lab values can be checked using the characteristics of the target. Such as, the XYZ values of the leftest patch of gray-bar in the target must larger than of the rightest patch of gray-bar in the target.

### **3.2. target-image-analyzer**

The target-image-analyzer is used to extract the RGB values of color patches from the IT8.7/1 or IT8.7/2 target image automatically. To extract the RGB values in the scanned target image, a target-positioning algorithm is proposed to determine the four-corners of the frame in the outer of the target image. After the four-corners of the frame are determined, the position of color patches can be derived by the posture of the color patches in the frame and the RGB values of color patches can be retrieved by averaging the RGB values of pixels inside the color patches. The target-positioning algorithm used to find the four-corners of the frame of the target image is listed in the following.

#### *The target-positioning algorithm*

- (1) Determine the position of the left and right borders of the frame: Firstly, the middle row of the target image is picked. To find the position of the left border in the middle row, the pixels are checked by border verifier from the most left pixel to the right until the border pixel is found. To find the position of the right border in the middle row, the pixels are checked by border verifier from the most right pixel to the left until the border pixel is found.
- (2) Determine the upper-left and upper-right corners of the frame: The left and right border pixels in the upper row of middle row can be found by checking the pixels in the top and near to the border pixels of the middle row with border verifier. By using the same procedure upgoing, all the left and right border pixels in the top of the left and right border pixels of the middle row can be found. As a result, the upper-left and upper-right corners of the frame can be found.
- (3) Determine the lower-left and lower-right corners of the frame: The left and right border pixels in the lower row of middle row can be found by checking the pixels in the bottom and near to the border pixels of the middle row with border verifier. By using the same procedure downgoing, all the left and right border pixels in the bottom of the left and right border pixels of the middle row can be found. As a result, the lower-left and lower-right corners of the frame can be found.

The illustration of the target-positioning algorithm is shown in figure 6.

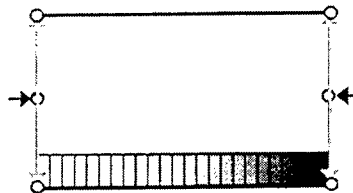


Figure 6. The process of the target-positioning algorithm

To recognize the border pixel, a border verifier is developed. The border verifier verifies the border pixels by using the border conditions and border weighting function. The border conditions and border weighting function are given in table 1. In step 1 of the target-positioning algorithm only the border conditions are used. In step 2 and step 3 of the target-positioning algorithm both the border conditions and border weighting function are used.

Table 1. The border condition table

border conditions	$Dl$	$Dr$	$Dlr$	$Wa$	$Wb$
(White, Black, Gray)	$>EdgeThr$	$>EdgeThr$	$>EdgeThr$	$PrivilegeW*3$	$Dl+Dr+Dlr$
(Gray, Black, Gray)	$>EdgeThr$	$>EdgeThr$	$<EdgeThr$	$PrivilegeW*2$	$Dl+Dr$
(Black, Black, Gray)	$<EdgeThr$	$>EdgeThr$	$>EdgeThr$	$PrivilegeW*1$	$Dlr$
(Gray, Black, White)	$>EdgeThr$	$>EdgeThr$	$>EdgeThr$	$PrivilegeW*3$	$Dl+Dr+Dlr$
(Gray, Black, Black)	$>EdgeThr$	$<EdgeThr$	$>EdgeThr$	$PrivilegeW*1$	$Dlr$
Others				0	0

The border conditions are observed from the relations of the border pixels and their neighboring pixels and denoted as the form of  $(L, B, R)$ . The  $L$  represents the pixel on the left of the border, the  $B$  represents the border pixel, and the  $R$  represents the pixel on the right of the border. The observed border conditions are: (White, Black, Gray), (Gray, Black, Gray), (Black, Black, Gray), (Gray, Black, White), and (Gray, Black, Black). For a given pixel, the  $Dl$ ,  $Dr$ ,  $Dlr$ , and  $EdgeThr$  are used to check if the pixel matches any border conditions. Where the  $Dl$  is the difference between the given pixel and its left neighboring pixel, the  $Dr$  is the difference between the given pixel and its right neighboring pixel, and the  $Dlr$  is the difference between the left neighboring pixel and the right neighboring pixel of the given pixel. The  $EdgeThr$  is a threshold and is used to determine the relation of  $LB$ ,  $BR$ , and  $LR$ .

The border weighting function  $BWF$  is defined to determine a best border pixel from plural possibly border pixels and listed in the following.

$$BWF = W_a + W_b \quad (5)$$

Where the  $W_a$  is observed from the difference between the border conditions and the  $W_b$  is observed from the difference between the border pixels belong to the same border condition. That is, the  $W_a$  is used to find the pixels of better border condition and the  $W_b$  is used to find the better border pixel of the same border condition. The  $PrivilegeW$  in table 1 is used to distinguish the effect of  $W_a$  and  $W_b$  and must larger than the maximum of all combination of  $Dl+Dr+Dlr$ .

### 3.3. profile-parameters-evaluator

The profile-parameters-evaluator is used to generate the parameters of the three-component matrix-based model and N-component LUT-based model for the scanner ICC Profile. The flowcharts of two models used for scanner ICC Profile are shown in figure 7 and figure 8.

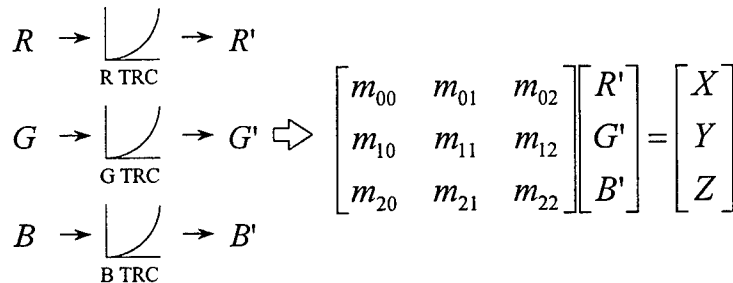


Figure 7. The three-component matrix-based model for scanner ICC Profile

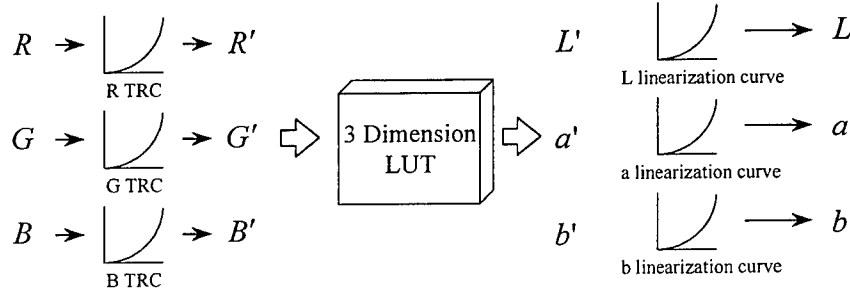


Figure 8. The N-component LUT-based model for scanner ICC Profile

For the three-component matrix-based model, the RGB TRCs and the RGB to XYZ transfer matrix (RGB colorants) must be generated. For the N-component LUT-based model, the RGB TRCs, the RGB to Lab LUT, and the linearization curves of Lab must be generated. The RGB TRCs used in the three-component matrix-based model are the same as the ones used in the N-component LUT-based model. The linearization curves of Lab are not used to linearize the Lab in this paper. That is, the L'a'b' values are equal to the Lab values. The generation of the RGB TRCs, the RGB to XYZ transfer matrix, and the RGB to Lab LUT are described as follows.

**The generation of RGB TRCs:** Figure 9 illustrates the generation method of the RGB TRCs in the implementation. Where the Y value and RGB values comes from the gray-bar of the target. The RGB TRCs are used to pull the RGB tone curves to fit the Y tone curve. By using the Y value and RGB values, the tone reproduction curve of RGB channels can be generated and can be used to linearize the RGB channels.

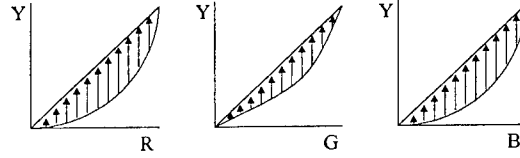


Figure 9. The generation of RGB TRCs

**The generation of the RGB to XYZ transfer matrix:** In this paper, the regression model [6-9] is used to create the RGB to XYZ transfer matrix. The color patches located in the column 1 to 19 of row A to L and the gray-scale 1 to 22 are used as the input of regression model. Before the RGB can be used, the RGB must be linearized using the RGB TRCs. For a set of linearized RGB and XYZ data  $\{R(l), G(l), B(l), X(l), Y(l), Z(l): l=1 \text{ to } n\}$ , the RGB to XYZ transfer matrix  $M$  can be derived using the following equations.

$$\begin{aligned} C(l) &= \{c_0(l), c_1(l), c_2(l)\} \\ &= \{R(l), G(l), B(l)\}; \end{aligned} \quad (6)$$

$$\begin{aligned} A_{3 \times 3} &= \{a(i, j): i = 0 \sim 2, j = 0 \sim 2\}, \\ a(i, j) &= \sum_{l=1}^n c_i(l) \cdot c_j(l); \end{aligned} \quad (7)$$

$$\begin{aligned} B_{3 \times 3} &= \{b(i, j): i = 0 \sim 2, j = 0 \sim 2\}, \\ b(i, 0) &= \sum_{l=1}^n X(l) \cdot c_i(l), \\ b(i, 1) &= \sum_{l=1}^n Y(l) \cdot c_i(l), \\ b(i, 2) &= \sum_{l=1}^n Z(l) \cdot c_i(l); \end{aligned} \quad (8)$$



$$M_{3 \times 3}^T = A_{3 \times 3}^{-1} \cdot B_{3 \times 3} \quad (9)$$

**The generation of the RGB to Lab LUT:** The generation process of the RGB to Lab LUT is shown in figure 10. In figure 10, the input RGB values with the size of  $33 \times 33 \times 33$  are firstly linearized by the RGB TRCs. Secondly, a  $3 \times 14$  matrix is used to translate the linearized RGB values to XYZ values. Finally, the Lab values of  $33 \times 33 \times 33$  LUT can be obtained by translating the XYZ values to Lab values using the XYZ to Lab function.

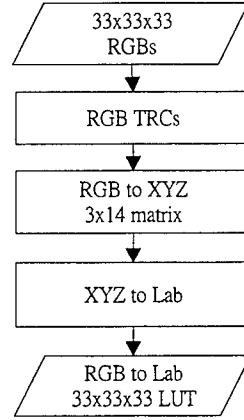


Figure 10. The generation process of the RGB to Lab LUT

The matrix used here is similar to the one presented in the [6] and is generated using the regression model. The generation process of the matrix is similar to the generation process of the RGB to XYZ transfer matrix. But sets the  $C(l) = \{c_0(l), c_1(l), c_2(l), c_3(l), c_4(l), c_5(l), c_6(l), c_7(l), c_8(l), c_9(l), c_{10}(l), c_{11}(l), c_{12}(l), c_{13}(l)\} = \{1, R(l), G(l), B(l), R(l)^2, G(l)^2, B(l)^2, R(l) \times G(l), G(l) \times B(l), R(l) \times B(l), R(l)^3, G(l)^3, B(l)^3, R(l) \times G(l) \times B(l)\}$ , the size of matrix A is  $14 \times 14$ , and the size of matrix B is  $14 \times 3$ .

#### 4. EXPERIMENTAL RESULTS

To demonstrate the performance of the scanner ICC Profile generator, a scanner ICC Profile evaluator is developed here to evaluate the performance of scanner ICC profile. The flowchart of the scanner ICC Profile evaluator is shown in the following.

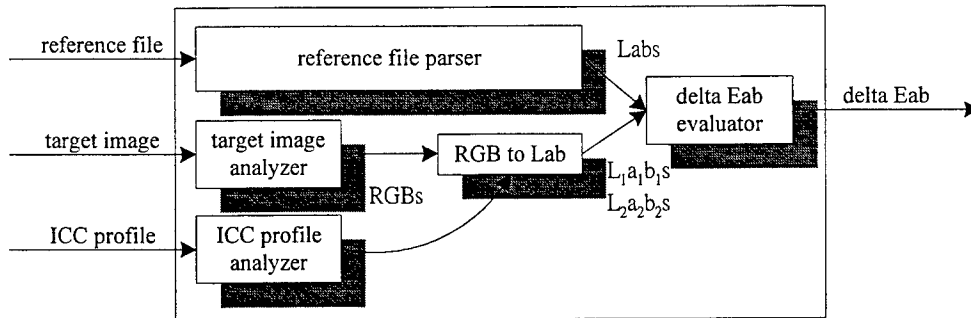


Figure 11. the flowchart of scanner ICC profile evaluator

The scanner ICC profile evaluator mainly consists of a reference-file-parser, a target-image-analyzer, an ICC-profile-analyzer, a RGB-to-Lab function, and a delta-Eab-evaluator. Where the reference-file-parser and the target image-analyzer are the same as of the scanner ICC profile generator. The ICC-profile-analyzer is used to retrieve the RGB TRCs, the RGB to XYZ transfer matrix, and the RGB to Lab LUT from the ICC profile. The RGB to Lab function translates the RGB values extracted from the target-image to two sets of Lab values for the three-component matrix-based model and N-component

LUT-based model, respectively. Where, a Trilinear interpolation scheme [10] is applied for the translation of RGB to Lab using the N-component LUT-based model. After the two sets of Lab values acquired, the  $\Delta E^*_{ab}$ s of the two models can be evaluated by the following equation.

$$\Delta E^*_{ab} = \sqrt{(L - L_1)^2 + (a - a_1)^2 + (b - b_1)^2} \quad (10)$$

In the experiments, a Kodak IT8.7/2 target produced in the march 1998 and four kinds of scanners - the Acer Prisa 620U, the HP SJ4200U, the Mustek 1200U, and the Umax Astra 2100U - are used to evaluate the performance of scanner ICC profile generator. In the experimental process, the ICC profiles of testing scanners are generated using the scanner ICC profile generator. Next, the same reference file and target images used to generate the ICC profiles are also used to evaluate the performance of the generated scanner ICC profile. The average  $\Delta E^*_{ab}$ s of tested scanners using the three-component matrix-based model and N-component LUT-based model are given in Table 2 and Table 3, respectively.

Table 2. The average  $\Delta E^*_{ab}$ s of four types of scanners using three-component matrix-based model

Scanner	$\Delta E^*_{ab}$
Acer Prisa 620U	3.48
HP SJ4200C	3.36
Mustek 1200U	4.09
Umax Astra 2100U	3.88

Table 3. The average  $\Delta E^*_{ab}$ s of four types of scanners using N-component LUT-based model

Scanner	$\Delta E^*_{ab}$
Acer Prisa 620U	2.86
HP SJ4200C	2.87
Mustek 1200U	2.85
Umax Astra 2100U	2.94

From the above tables, we can see the average  $\Delta E^*_{ab}$ s of tested scanners are almost less than 4 in three-component matrix-based model and the average  $\Delta E^*_{ab}$ s of tested scanners are less than 3 in N-component LUT-based model.

## 5. CONCLUSIONS

In this paper, an implementation of scanner ICC profile generator is described. A reference-file-parser and a target-image-analyzer are developed to automate the retrieving of the XYZ and Lab values from the reference file and the extraction of the RGB values of color patches from the target image, respectively. Two regression models are used to implement the three-component matrix-based model and N-component LUT-based model. The experimental results show the average  $\Delta E^*_{ab}$ s of generated ICC profiles for both models are small.

## REFERENCES

1. International Color Consortium, "ICC.1," Sep., 1998.
2. International Color Consortium, "ICC.1A," April, 1999.
3. ANSI IT8.7/1, "Graphic Technology - Color transmission target for input scanner calibration," 1993.
4. ANSI IT8.7/2, "Graphic Technology - Color reflection target for input scanner calibration," 1993.
5. ANSI CGATS5, "Graphic technology - Spectral measurement and colorimetric computation for graphic arts images," 1993.
6. Henry R. Kang, "Color scanner calibration," J. Imaging Technol., Vol. 36, No. 2, pp. 162-170, March 1992.
7. Shoji Suzuki, Tadakazu Kusunoki, and Masahiro Mori, "Color characteristic design for color scanners," Applied Optics, Vol. 29, No. 34, pp. 5187-5192, Dec. 1990.
8. Po-Chieh Hung, "Colorimetric calibration for scanners and media," Proc. SPIE, vol 1448, pp. 164-174, 1991.
9. A. A. Afifi and S. P. Azen, Statistical Analysis, Academic Press, NY, 1972.
10. J. M. Kasson, "Performing color space conversions with three-dimension linear interpolation," J. Electronic Imaging, 4(3), pp. 226-250, 1995.

# Color reproduction system based on color appearance model and gamut mapping

Fang-Hsuan Cheng, Chih-Yuan Yang

Department of Computer Science & Information Engineering, Chung Hua University  
Hsin-Chu 300, Taiwan, ROC

## ABSTRACT

By the progress of computer, computer peripherals such as color monitor and printer are often used to generate color image. However, cross media color reproduction by human perception is usually different. Basically, the influence factors are device calibration and characterization, viewing condition, device gamut and human psychology. In this thesis, a color reproduction system based on color appearance model and gamut mapping is proposed. It consists of four parts; device characterization, color management technique, color appearance model and gamut mapping.

**Keywords:** Color appearance model, Gamut mapping

## 1. INTRODUCTION

With the recent advent of color management systems, there provide predictable and consistent color results between different imaging peripherals. We can use them to get more satisfactory color quality than past. Because color is pervasive across media, some unresolved issues have produced. The ICC Processing Model assumes that the Profile Connection Space represents a perfect reproduction. It consists of the ideal reference viewing conditions, perfectly reflecting and unlimited gamut of colorants. In practical application of cross media reproductions, a reproduction will look exactly the same as original image only if both have the same XYZ values for the white point, the two media have similar surface characteristics and are observed under similar viewing conditions, and the reproduction medium can produce all the colors present in the original. So it is important to consider these factors to get good reproduction results.

For these reasons above, we want to develop a color reproduction system that can make sure the color consistent and provide predictable color results between different media. In addition, we hope our application can reproduce more consistent image without measuring instruments. So we use Microsoft Image Color Management (ICM) technique to solve media differences, without using any measuring instruments.

As far as color reproduction is concerned, we may take a deep insight of this problem by dividing it into several parts. They are described as followings:

### 1.1 Device calibration and characterization

Device calibration is the setting of the imaging device to a known state. Calibration ensures that the device is producing consistent results, both from day to day and from device to device. Device characterization defines the relationship between the device color space and the CIE system of color measurement [1]. There are three main approaches to device characterization: physical modeling, empirical modeling and exhaustive measurement. For physical modeling of imaging devices, it involves building mathematical models that relate the calorimetric coordinates of the input or output image elements to the signals used to drive an output device or the signals originating from an input device. For empirical modeling of imaging devices, it involves collecting a fairly large set of data and then statistically fitting a relationship between device coordinates and calorimetric coordinates. For exhaustive measurement of imaging devices, it involves exhaustive measurement of the output for a complete sampling of the device's gamut. But it has a disadvantage that large number of measurements must be made.

Different types of calorimetric measurements are required for the characterization of various imaging devices. For CRT monitor, an overview of alternative display technologies can be found by Jackson [2] and Budin [3]. Beside Berns [4] provide further details on the measurement and characterization of CRT displays. For scanner and digital cameras, the colorimetric calibration and characterization of input devices have been described by Rodriguez and Stockham [5]. For

printers and other output devices, Yule and Nielsen [6] proposed the modified version of Neugebauer Equation with using a LUT between the digital values and measured density.

## 1.2 Color Appearance Model

In order to solved the problem of mismatch between the output as a printer and that of a monitor. Before 1990, the work of color reproduction focused on the consistency of calorimetric measurements between different media [7,8]. However, this type of research did not consider the effect of illumination and can not solve the problem of cross-media image reproduction completely. At the same time, some other researches demonstrated that the appearance of color can be affected by not only the color stimulus but also its viewing conditions like reference white, luminance, surrounding and background [9,10,11]. Until now, several color appearance models had been proposed [12,13,14] to predict color appearance for specific conditions. The aim of these color appearance models is to provide consistency and predictable appearance match between different media.

## 1.3 Gamut Mapping

Different devices are capable of producing different range of colors. The range of colors associated with a device is known as its gamut. Gamut mapping is perhaps the most important element in transforming images across media. It is a fairly new topic in the literature.

Stone, Cowan and Beaty [15] investigated both clipping and compressing techniques in XYZ space. Gentile, Walowitt and Allebach [16] compared several methods of gamut mismatch compensation in  $Lu^*v^*$  space. Generally, they preferred clipping in chroma while keeping lightness and hue constant. Parsier [17] performed a study similar to that of Gentle but with hard copy images. He found that, depending on the input image, either of two techniques was preferred. The first preserves lightness and hue while clipping chroma. The second preserves hue while clipping lightness and chroma toward to the (50, 0, 0) point of  $La^*b^*$  space. MacDonald [18] investigated various mapping in the Hunt [19] color appearance space. His preferred method is also simultaneous compression of light and chroma toward a mid-gamut point. Hoshino and Berns [20] looked at lightness mappings in the Hunt color appearance space. They introduced the concept of "soft compression" in which a cut-off point is defined on the axis of interest. Compression takes place only for values above the cut-off. Wolski, Allebach and Bouman [21] investigated three mapping methods which preserve hue while changing lightness and saturation. First, they altered saturation only. Second, they changed lightness only. Third, they simultaneously clip lightness and saturation toward the center of the target gamut. A linear map, proposed by Lamming and Rhodes [22], which to scale and translate the monitor  $L^*$  to the printer range for  $L^*$  is adequate if the image already appears satisfactorily on the monitor. Meyer and Barth [23] suggested a local function that both adapts the lightness scale and provides edge enhancement. Stone and Wallace [24] proposed an approach to non-linearly adjust the image colors in lightness to control the dynamic range, and in chroma to bring overlay saturated colors inside the target gamut. Most gamut mapping methods do not adjust the hue angle. But the gamut mapping method of Ruetz and Brunoe [25] warps hue angle to compensate for Abney effect [25].

## 2. Proposed approach

The proposed color reproduction system can be divided into three parts; color management technique, color appearance model and gamut mapping. Since the color space representation of each device is different, we must transform the different color space into a device independent color space. This is so-called color space transformation. In order to get more accuracy XYZ values, we using Microsoft ICM 2.0 API functions to transform color space between different device. In addition, the Microsoft ICM 2.0 API functions also control the input and output of different devices. After transformation, the color appearance model will be performed. The XYZ data are transformed via the color appearance model into perceptual LCH coordinates, using the parameters that define the monitor viewing conditions. For gamut mapping model, the LCH image will be modified by compressing the colors that are outside the printer gamut onto the boundary of the gamut. After gamut mapping model, the modified LCH image is then transformed via the inverse color appearance model back into XYZ, using the parameters that define target print viewing conditions. Finally, the XYZ image is converted to CMYK ink values and printed out via printer profile using the Microsoft ICM 2.0 API functions. According to these processes, we can compare the image that printed out with that showed on monitor.

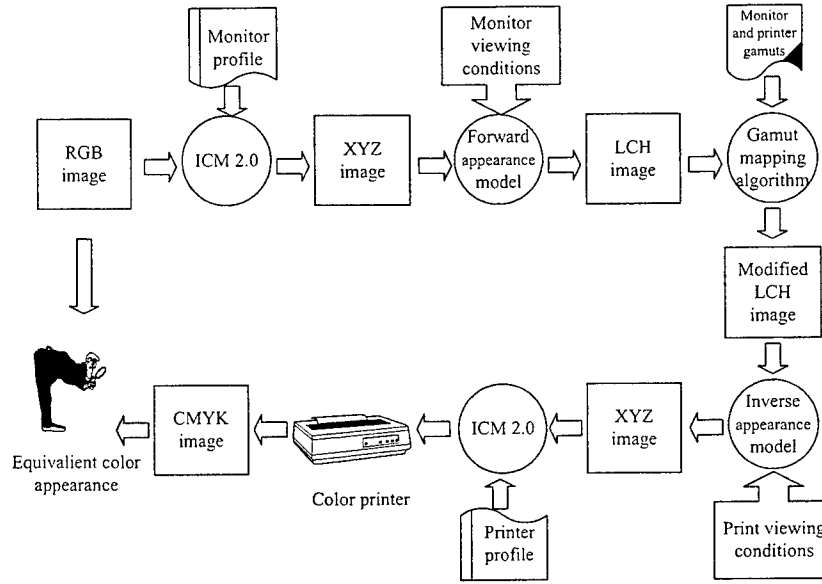


Figure 1. System overview

### 2.1. RLAB model

RLAB model was proposed by Fairchild and Berns [27] for cross media image reproduction application. It evolved from studies of chromatic adaptation; fundamental CIE colorimetry and practical implications in cross media image reproduction. The concept of RLAB is to take advantage of the good spacing under daylight and familiarity of the CIELAB space, while improving its applicability to non daylight illuminants. It can be used to calculate correlates of lightness, chroma, saturation and hue, but it can not be used to predict brightness or colorfulness. For input data of the RLAB model include the relative tristimulus values of the test stimulus (XYZ) and the white point ( $X_n, Y_n, Z_n$ ), the absolute luminance of a white object in the scene, the relative luminance of the surround (dark, dim, average) and a decision on whether discounting-the-illuminant is taking place. The forward implementation of RLAB model is described as following. One begins with a conversion from CIE tristimulus values ( $Y=100$  for white) to fundamental tristimulus values with

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = M \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad M = \begin{bmatrix} 0.3897 & 0.6890 & -0.0787 \\ -0.2298 & 1.1834 & 0.0464 \\ 0.0 & 0.0 & 1.0000 \end{bmatrix} \quad (1)$$

The next step is the calculation of the A matrix that is used to model the chromatic-adaptation transformation. The A matrix represents von Kries adaptation coefficients that are applied to the cone responses the test stimulus (LMS). The A matrix can be calculated by following Equations.

$$A = \begin{bmatrix} a_L & 0.0 & 0.0 \\ 0.0 & a_M & 0.0 \\ 0.0 & 0.0 & a_S \end{bmatrix} \quad a_L = \frac{P_L + D(1.0 - p_L)}{L_n} \quad a_M = \frac{p_M + D(1.0 - p_M)}{M_n} \quad a_S = \frac{p_S + D(1.0 - p_S)}{S_n} \quad (2)$$

Where  $Y_n$  is the absolute adapting luminance. The cone response terms with  $n$  subscripts ( $L_n$ ,  $M_n$ ,  $S_n$ ) refer to values for the adapting stimulus derived from relative tristimulus values. The  $D$  factor allows various proportions of cognitive discounting-the-illuminant. The adjustable  $D$  parameter for media in RLAB model shown in Table 1.

$$p_L = \frac{(1.0 + Y_n^{1/3} + l_E)}{(1.0 + Y_n^{1/3} + 1.0/l_E)} \quad p_M = \frac{(1.0 + Y_n^{1/3} + m_E)}{(1.0 + Y_n^{1/3} + 1.0/m_E)} \quad p_S = \frac{(1.0 + Y_n^{1/3} + s_E)}{(1.0 + Y_n^{1/3} + 1.0/s_E)} \quad (3)$$

**Table 1.** The adjustable  $D$  parameter for media in RLAB model

Media	Parameter value
Hardcopy image	1.0
Softcopy image	0.0
Projected transparencies	0.5

$$l_E = \frac{3.0L_n}{L_n + M_n + S_n} \quad m_E = \frac{3.0M_n}{L_n + M_n + S_n} \quad s_E = \frac{3.0S_n}{L_n + M_n + S_n} \quad (4)$$

After the  $A$  matrix is calculated, the tristimulus values for a stimulus color are converted to corresponding tristimulus values under the reference viewing conditions can be obtained by following Equations.

The RLAB coordinates are then calculated using the following Equations.

$$\begin{bmatrix} X_{ref} \\ Y_{ref} \\ Z_{ref} \end{bmatrix} = RAM \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad R = \begin{bmatrix} 1.9569 & -1.1882 & 0.2313 \\ 0.3612 & 0.6388 & 0.0 \\ 0.0 & 0.0 & 1.0000 \end{bmatrix}$$

$$\begin{aligned} L_R &= 100(Y_{ref})^\sigma \\ a_R &= 430[(X_{ref})^\sigma - (Y_{ref})^\sigma] \\ b_R &= 170[(Y_{ref})^\sigma - (Z_{ref})^\sigma] \end{aligned} \quad (5)$$

$L_R$  represents an achromatic response analogous to CIE  $L^*$ . The red-green chromatic response is given by  $a_R$  (analogous to CIELAB  $a^*$ ) and yellow-blue chromatic response is given by  $b_R$  (analogous to CIELAB  $b^*$ ). For the input parameter  $\sigma$ , it represents the relative luminance of the surround. Its corresponding values are shown in Table 2.

**Table 2.** The adjustable  $\sigma$  parameter for surrounds in RLAB model

Surrounds	Parameter value
Average surrounds	1/2.3
Dim surrounds	1/2.9
Dark surrounds	1/3.5

## 2.2. Finding gamut of printer

Before doing gamut mapping, we must know the gamut of printer. So we print the patches by EPSON stylus color printer and use the standard of ISO 12642 [9] that defines a color palette consisting of 928 combinations of cyan, magenta, yellow and black ink values. After printing, we get the  $L^*a^*b^*$  values of patches by X-Rite 938 spectrophotometer.

Two sets of ink values are specified which span, with differing intervals, the color space defined by combinations of cyan, magenta, yellow and black dot area percentages. The basic data set, which is a subset of the extended data set, shall be the default set in the absence of any other information; the extended data set (or subsets of the ink value data set) may be used if specified. The data are defined as digital data and are not the printed image values (or sets of separations). However, the colorimetric values needed to produce the color characterization data file may be determined by printing images which have been made from films containing halftone values corresponding to the values in the ink value data set. For example, it can be mapped the value from 100 to 255 for four colors individual.

### 2.2.1 Mapping Lightness

The  $L^*$  value represented the lightness in the  $L^*a^*b^*$  system. In absolute terms, monitors are much more dim than printed pages under normal viewing conditions. The darkest black printable may be lighter (under some given illumination) than the brightest light given off by a monitor. Therefore, it is not the absolute lightness but some relative measure that needs to be considered. In the  $L^*a^*b^*$  system, lightness is measured relative to the brightest achromatic color. When transformed to  $L^*$  values, the point corresponding to "white" is always  $L^*=95$ . However, the black point is in different position relative to the white point on different devices. Typical values for monitor black are  $L^*=2$  or 3. For a printer black, we set the values as high as  $L^*=35$  according our experiment.

$L^*$  axis is linearly scaled and transformed so that the  $L^*$  value for black on input is equal to or slightly less than the minimum  $L^*$  value that is black on output. Using a value less than the true black value means that the darkest colors are projected to a point on surface of the destination gamut. This will produce images with improved contrast compared to exactly matching the black values, at the cost of detail in the dark regions. Image that is not very dark tends not to lose much detail using this method. With very dark images, the black point should be matched exactly so that minimum detail is lost. Note that a printer's gamut is much more narrow around the black point than the monitor's gamut, so some compression of the colors in the dark area is inevitable.

### 2.2.2 Mapping Hue Angle

In most of gamut mapping methods, none of them change the hue angle. More particularly, region of pure yellow colors for the printer falls into a very narrow range of the printer gamut. The range of monitor yellow colors is greater. Because the range of pure yellow for printer colors is so narrow that a user typically obtain a greenish-yellow rather than the desired pure yellow colors. Thus the yellow region is widened. Conveniently, yellow widening is obtained through hue angle warping [28] as follows :

$$\begin{aligned} &\text{For hue angle between } 87^\circ \text{ to } 91^\circ: \\ &\quad \text{warped angle} = 87 + 1.25 * (\text{ang} - 87) \end{aligned} \tag{6}$$

$$\begin{aligned} &\text{For hue angle between } 97^\circ \text{ to } 112^\circ: \\ &\quad \text{warped angle} = 92 + 0.5 * (\text{ang} - 97) \end{aligned} \tag{7}$$

$$\begin{aligned} &\text{For hue angle between } 97^\circ \text{ to } 112^\circ: \\ &\quad \text{warped angle} = 92 + 0.5 * (\text{ang} - 97) \end{aligned} \tag{8}$$

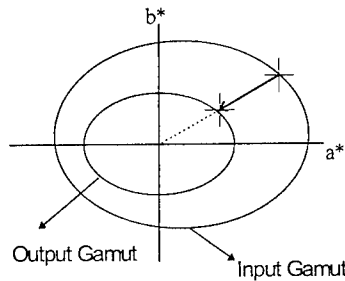
$$\begin{aligned} &\text{For hue angle between } 112^\circ \text{ to } 132^\circ: \\ &\quad \text{warped angle} = 99.5 + 1.25 * (\text{ang} - 112) \end{aligned} \tag{9}$$

$$\begin{aligned} &\text{For hue angle between } 132^\circ \text{ to } 147^\circ: \\ &\text{warped angle} = 124.5 + 1.5 * (\text{ang} - 132) \end{aligned} \quad (10)$$

where ang is the hue angle

### 2.2.3 Mapping Chroma

The simplest form of chroma compression projects all out-of-gamut values to the surface of the target gamut. So we replaced the out-of-gamut point with the point that is nearest it and on the border of the curve of gamut. This method is shown in Figure 2. Compressing chroma along lines of constant hue angle provides a method for mapping such an image gamut into a printer gamut without changing the hue, within the limits of the hue definition for uniform color space.



**Figure 2.** The method of mapping chroma

## 3. Experiment results

The printer used in this experiment is EPSON Stylus COLOR Ink jet printer. Although the maximum resolution of the printer is 720 dpi, only 360 dpi is used to implement the algorithm. We use the standard of ISO 12642 that defines a color palette consisting of 928 combinations of cyan, magenta, yellow and black ink values. Therefore, 928 color patches are printed and then measured by spectrophotometer of X-Rite 938 under  $D_{65}$  illuminant. The data are recorded as CIELAB values. The results were shown in Figure 3, 4 and 5.

The datum we measured is used to find the boundary of printer gamut. The data set can be used to do gamut mapping. Besides, we need to print the nine reference colors of cyan, magenta, yellow, red, green, blue, mixed-color (CMY), black and white (no ink), and measure the actual CIELAB values for these colors.

For the color appearance model, we use the RLAB model in our system. The experiment is conducted in a dim surrounding. Since the aim of our method is want the reproduced images with the original images on monitor more closely. In RLAB model, we set the D factor equal to 1.0 for hardcopy images and 0.0 for softcopy images.

Finally, Reproduced images by printer are compared with original images that displayed on NEC multiSync 5FGp CRT. Twenty observers took part in the experiment. Most observers are in the field of image processing. They ranged in age from 23 to 28 years old. We use the IT8 standard graphics images to test our algorithm. Each image will be printed by three different methods including our proposed method.



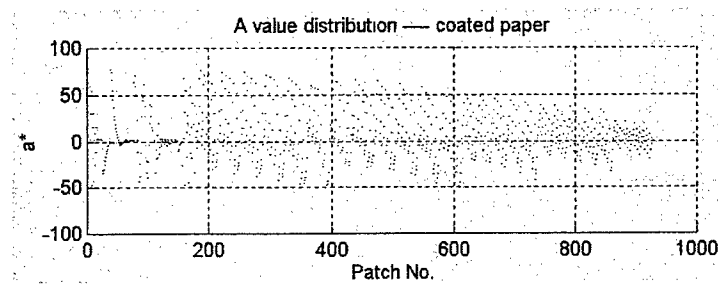


Figure 2. The value of  $a^*$

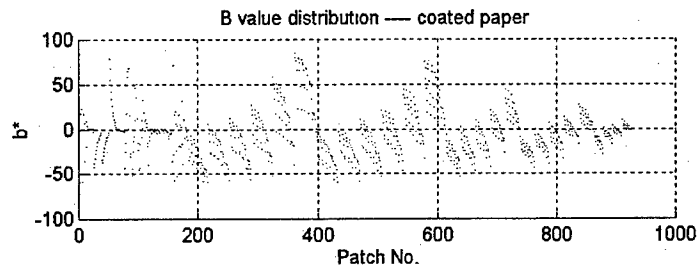


Figure 3. The value of  $b^*$

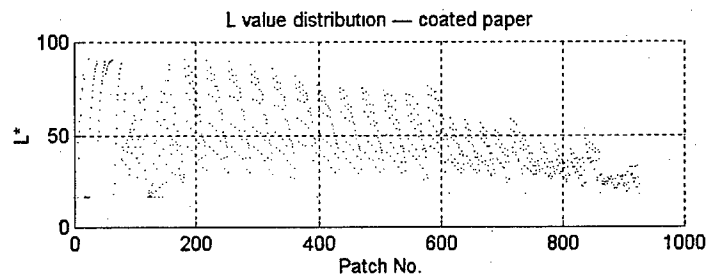


Figure 4. The value of  $L^*$

### 3.2 Comparison of images

In our experiments, three ANSI IT8 standard images are used as the tested images. They consist of format with true color. The first is gloomy, and the contrast of color between light and dark is strong. The color of second image is rich, and it can test color reproduction of our system. The third image consists of most light colors. It is tested for burnish of the metal and gray color rendering. The contents and characteristics of the testing images shown in Table 3.

Moreover, all images are printed by using different four kinds of method. Because we can not present the same visual effect for tested images on monitor as the images on papers. First, we use Fujix pictography 3000 to print the images as the reference image. Second, the tested images are printed by Kodak Imaging. Third, we printed the tested images by Kodak Imaging with using ICM 2.0. Finally, we use the proposed algorithm to print the images.

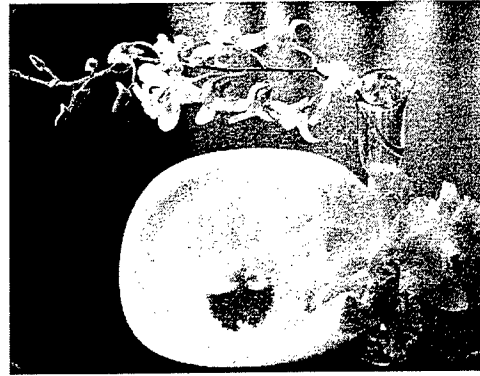
The part (a) of each image is printed at 400 dpi by Fujix pictography 3000. The part (b) of each image is printed by Kodak Imaging. The part (c) of each image is printed by Kodak Imaging with using ICM 2.0. The part (d) of each image is printed by our method.

**Table 3.** The contents and characteristics of the testing images

	contents	characteristics
Image #1	orchid	Low Lightness
Image #2	fruits	High saturation
Image #3	tableware	High Lightness



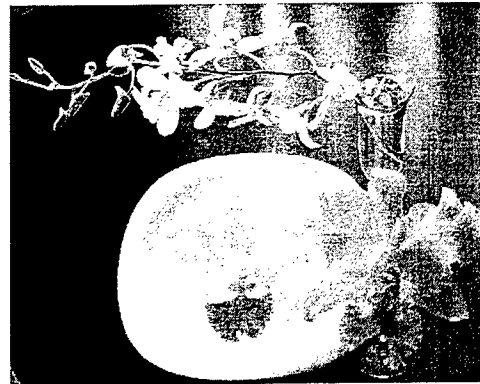
(a)



(b)



(c)

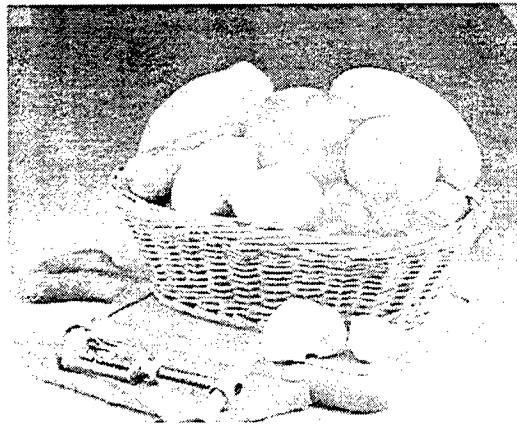


(d)

**Figure 5.** Sample image #1



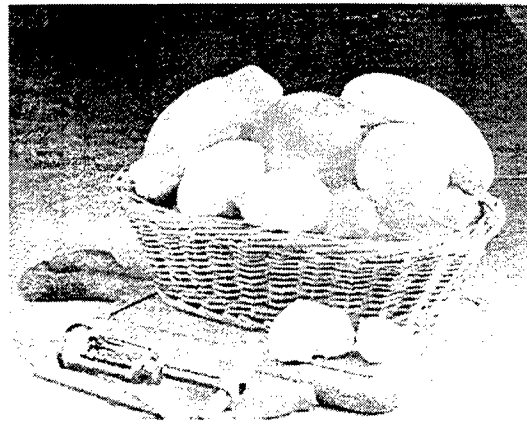
(a)



(b)



(c)



(d)

Figure 6. Sample image #2

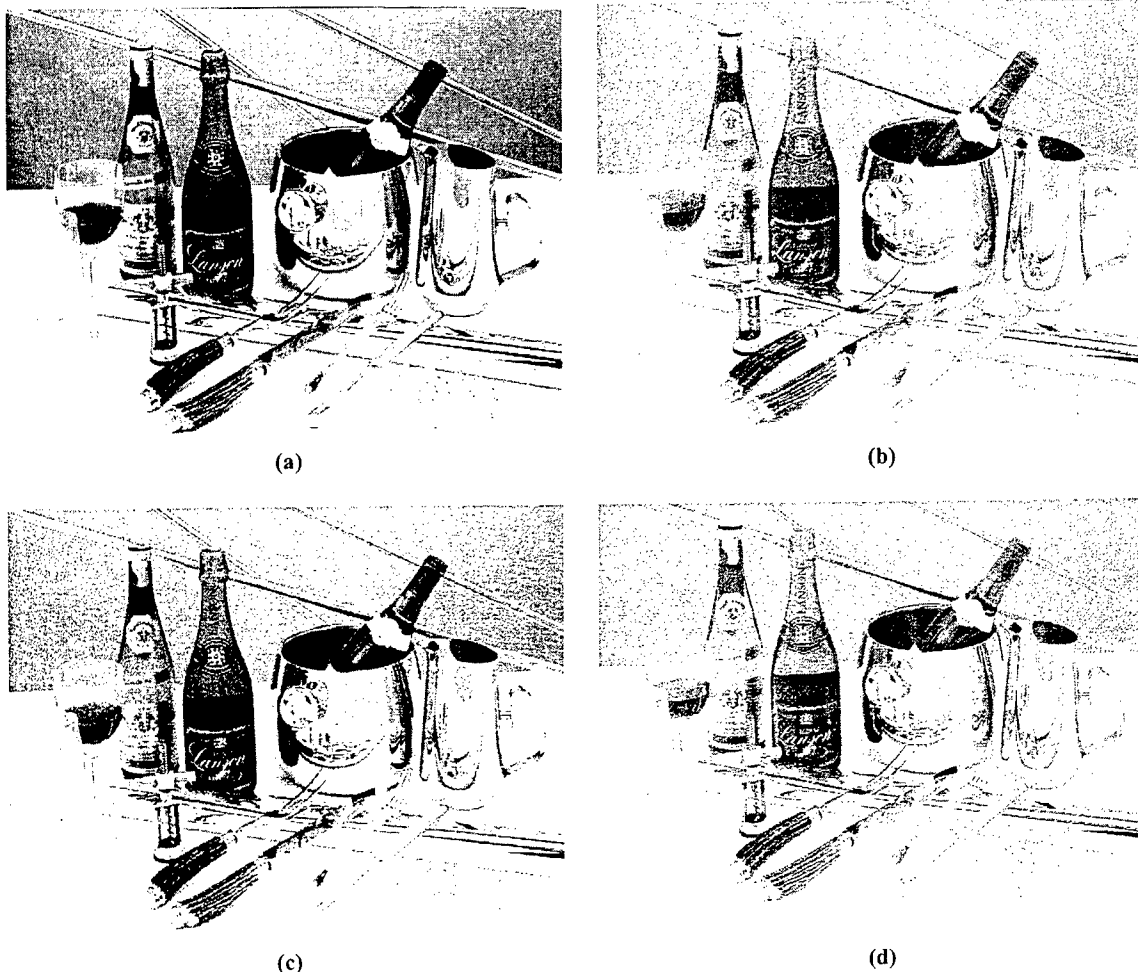


Figure 7. Sample image #3

Taking a global view of these three images, the color is more like the display of monitor when they were printed by our method. The result has shown in Table 4. For the tested image #1, There are almost seventy-percent observers to decide that printed by our method is more consistent than others are. For the tested image #2, there are only forty observers to decide our method is better than others are. For the tested image #3, most of observers to decide our method have best color reproduction. For all of the tested images, no body to decide that printed by Imaging is better than others.

Table 4. Comparison result

	Printed by Imaging	Printed by Imaging with using ICM	Printed by our method
Image #1	0	6	14
Image #2	0	12	8
Image #3	0	5	15

#### 4. Conclusion

Since the computer peripherals become cheaper and more popular, it is important for us to improve consistence of cross media color reproduction. However, effective color reproduction remains a very complex process and depends on many factors, not only device behavior but also human visual perception and the viewing conditions in which an image is seen.

In our experiments, the color images implemented by our method have better color reproduction. Besides, our method can provide most of people to get consistent and accurate color reproduction without using "high end" peripherals and measured instruments.

There are still some the future work to be done. Since color appearance models are still the topic of research for color science. How to find the best performance of color appearance models in considering practical viewing conditions is yet to achieve. Second, if the better model of gamut mapping is found, the color reproduction system will be more perfect.

#### References

- [1] John A. J. (1996) "Methods for Characterising Colour Scanners and Digital Cameras", Displays Special Issue: 'To Achieve WYSIWYG Colours', 183-192
- [2] Jackson R., MacDonald L. W. and Freeman K. (1994) Computer Generated Color, John Wiley & Sons, New York
- [3] Budin J.P. (1995) "Emissive Displays: the Relative Merits of ACTFEL, Plasma and FEDs, Getting The Best from State-of-the-Art Display System". SID Conference Proceedings, London
- [4] Berns R. S. (1996) "Methods for Characterising CRT displays, Displays". Special Issue: 'To Achieve WYSIWYG Color', 173-182
- [5] M. A. Rodriguez and T. G. Stockham, "Producing calorimetric data from densitometric scans" Proc. SPIE 1913, 413-418 (1999)
- [6] Yule J. A. C. and Nielsen W.J. (1951) "The Penetration of Light into Paper and its Effect on Halftone Reproduction", TAGA Proceedings, 65-76
- [7] R. W. G. Hunt, "The Reproduction of Color in Photography, Printing, & Television". FOUNTAIN PRESS, 4 ed., 1987.
- [8] W. S. Stiles and G. Wyszecki, Color Science. New York: Wiley, 1982
- [9] ISO CD 12642, Draft of Sept 19, 1993.
- [10] H. E. J. Neugebauer, "Die Theoretischen Grundlagen des Mehrfarbendruckes" Z.wiss. Photogr., 36, 73-89, 1937.
- [11] R. W. G. Hunt, "Measuring Colour" 2nd Ed. Ellis Horwood, England, 1992.
- [12] R.G.W. Hunt, "The Reproduction of Colour in Photography, Printing & Television" 4th Ed. Fountain Press, England, 1987.
- [13] G. Wyszecki and W. S. Stiles, "Color Science : Concepts and Methods, Quantitative Data and Formula," 2nd Ed. John Wiley & Sons, New York, 1982.
- [14] Takebaru Uchizono, image Processing Apparatus, US Patent no. 5200839, Apr. 6, 1993.
- [15] M.C. Stone, W.B. Cowan and J.C. Beatty, "Color Gamut Mapping and the Printing of Digital Color Images" ACM Tran. on Graphics Vol. 7, No. 4, Oct. 1988.
- [16] R. S. Gentile, E. Waloeitt and J. P. Allebach, "A Comparison of Techniques for Color Gamut Mapping Mismatch Compensation" Journal of Imaging Tech., 16, 176-181, 1990.
- [17] E. G. Pariser, "An Investigation of Color Gamut Reduction Techniques" IS&T Symposium on Electronic Prepress Technology - Color Printing, 105-107, 1991.
- [18] L. W. MacDonald, "Gamut Mapping in Perceptual Color Space" Proc. IS&T/SID Color Imaging Conference : Transforms and Transportability of Color 1, 193-196, 1993.
- [19] R. W. G. Hunt, "Revised Colour-Appearance Model for Related and Unrelated Colours" Color Research and Application, 16, 1991.
- [20] T. Hoshino and R. S. Berns. "Color Gamut Mapping Techniques for Color Hard Copy Images" Device Independent Color Imaging and Imaging Systems Intergration, SPIE Vol. 1909, 152-164, 1993.
- [21] M. Wolski, J. P. Allebach and C. A. Bouman, "Gamut Mapping : Squeezing the Most Out of Your Color System" IS&T and SID's 2nd Color Imaging Conference, 1994.
- [22] M. J. Lamming and W. S. Rhodes, "A Simple Meth for Improved Color Printing of Monitor Images" ACM Transactions on Graphics, 94, p345-375, 1990.
- [23] J. Meyer, B. Barth, "Color Gamut Mapping for Hard Copy" in the Proceedings of the Optical Society of America

- Meeting on Applied Color Vision, July 12-14, p138-143, 1989.
- [24] M. C. Stone and W. E. Wallace, "*Gamut Mapping Computer Generated Imagery*" Graphics Interface 1991.
  - [25] B. Ruetz and S. Bruno, "*Color Printing methos and Apparatus Using an Out-of-Gamut Color Table*" United States Patent 5,299,291, Mar., 1994.
  - [26] Microsoft, "*An Overview of Microsoft Image Color Management Technology*".
  - [27] M. D. Fairchild and R. S. Berns, "*Image color-appearance specification through extension of CIELAB*" Color research and application, vol.18, pp.178- 190, June 1993.
  - [28] B. Ruetz and S. Bruno, "*Color Printing methos and Apparatus Using an Out-of-Gamut Color Table*" United States Patent 5,299,291, Mar., 1994.

## Poster Session

# Characteristic Extraction of Face using DWT and Recognition Based on Neural Networks

Hyung-Bum Kim<sup>a</sup>, Chun-Hwan Lim<sup>c</sup>, Seung-Jin Park<sup>b</sup>, Jong-An Park<sup>\*a</sup>

<sup>c</sup>Dept. of Electronic Information Eng, Dongkang College, Kwangju, Republic of Korea.

<sup>b</sup>Dept of Biomedical Eng., Chunnam Univ., Kwangju, Republic of Korea.

<sup>a</sup>School of Electronic Information & Communication Eng., Chosun Univ.

#375 Seosuk-Dong, Kwangju, Republic of Korea.

## ABSTRACT

In this paper, we suggest how to segment the face when there is the man under complex environment, extract the features from segmented the image and propose an effective recognition system using the discrete wavelet transform (DWT). This algorithm is proposed by following processes. First, two gray-level images are captured with 256 levels of the size of  $256 \times 256$  in constant illumination. We use a Gaussian filter to remove noise of input image and get a differential image between background and input image. Second, a mask is made from erosion and dilation process after binary of the differential image. Third, facial image is divided by projecting the mask into input image. Most characteristic information of human face is in eyebrow, eyes, nose and mouth. In the reason, the facial characteristic are detected after selecting local area including this area. Fourth, detecting the characteristic of segmented face image, edge is detected with Sobel operator. Then, eye area and the center of face are searched by using horizontal and vertical components of edge. Characteristic area consists of eyes, a nose, a mouth, eye brows and cheeks, is detected by searching the edge of the image. Finally, characteristic vectors are extracted from performing DWT of this characteristic area and are normalized it between +1 and -1. Normalized vectors are used with input vector of neural network. Simulation results show recognition rate of 100 % about learned image and 92% about test image.

**Keyword :** Mask, Differential image, Discrete wavelet transform, Neural network, Face recognition

## 1. INTRODUCTION

Edge image is mainly used to discriminate whether there is face or not in the image. Sakai applied oval mask to edge map extracted from input image, set the approximate head area, checked the edge image of eyes and mouth within the head area and then extracted the final head area<sup>[1]</sup>. This method has the weakness that it is influenced greatly by the direction of lighting. Kelly also produced the downward image interpretation method that extracts the outline of head and body automatically from input image and continuously extract the location of eyes, nose and mouth<sup>[2]</sup>. Craw and others proposed the method extracting head area using mask of hierarchical size in a given image<sup>[3]</sup>. Using the outline of head

---

\* Correspondence : Email: [japark@mail.chosun.ac.kr](mailto:japark@mail.chosun.ac.kr); Telephone:082-62-230-7064; Fax:082-62-232-6776



which is composed of edge image, Govindaraju searched the face in the complex background image<sup>[4]</sup>. Sirohey segmented the face by using edge image and brightness image which are extracted with Canny edge searcher in the image with background<sup>[5]</sup>. This method showed the precision of about 80% about 48 images without any restriction. In case of recognizing segmented images, feature extraction methods consists of local feature extraction and global feature extraction algorithm.<sup>[6]</sup> Feature location detection methods include the method by geometrical symmetry, method using correlation of feature templates like eyes, nose and mouth and image, method detecting face to search face candidate edges with snakelets, method detecting features of face with self-organizing feature map and method extracting the feature from frequency domain like FFT, DCT. After extracting the features, generally, Euclidean distance and neural network is used to recognize the face. While the former has the strength that system implementation is easy, it has also weakness that recognition rate is lowered when there are a lot of data. In case of performing pattern recognition with the latter, Input images are converted into spatial or frequency domain. Then, feature parameters of the image are extracted and it is used with input vectors of neural network. This method has the properties that the number of nodes and connection lines can be reduced and system implementation are easy<sup>[7]</sup>. In this paper, we propose that use differential image to obtain a mask and segment the face using a mask. Also, we define the characteristic areas of face to reduce the computational complexity in extracting the characteristic. And we extract the exact characteristic vectors using discrete wavelet conversion. This method reduces the number of characteristic vectors and decreases the required learning data in learning with neural network. In section 2, we describe face segmentation algorithm based on differential image. Section 3 describes face recognition algorithm based on wavelet. Section 4 describe simulation results. Conclusions are given in section 5.

## 2. FACE SEGMENTATION ALGORITHM BASED ON DIFFERENTIAL IMAGE

### 2.1 Face Segmentation Using Differential Image and Mask

The face is segmented by applying differential images in regular illumination condition. Image is obtained with gray scale 256 level of the size of 256×256 and the noise in the image is rejected by Gaussian filtering. And since the pixel value outside the face in input image including the face is not identical with that of background image which doesn't include the face exactly in obtaining the difference of input images including background image and face, threshold is given as shown in the following Expression (1) and the pixel value of the image obtained with the same camera can get the desired image though it is changed a little.

$$\begin{aligned}
 &\text{If } |Image1(x,y) - Image2(x,y)| < threshold \\
 &\quad \text{Then } Differ\_Image(x,y) = 0 \\
 &\text{Else } Differ\_Image(x,y) = |Image1(x,y) - Image2(x,y)|
 \end{aligned} \tag{1}$$

For, *Image1*, *Image2* : input image, *Differ\_Image* : differential image

The noise occurred by the change of illumination and the reflection of light occupies small area in the image because of low frequency of noise and it makes binary differential image and erodes the image. Eroded binary image should be dilated to the size of face image. Then pixel value is examined, mask is made, it is projected on the original image with the face and

the face is segmented in the background. Fig. 1 shows the flow chart to segment the face.

## 2.2 Feature Detection of Face

Most feature information of human face is in eyebrow, eyes, nose, mouth and cheek. So the face features are detected after selecting local area including this area. For the feature detection of segmented face image, edge is detected with sobel operator and eye area and the center of face are searched by using horizontal and vertical components of edge.

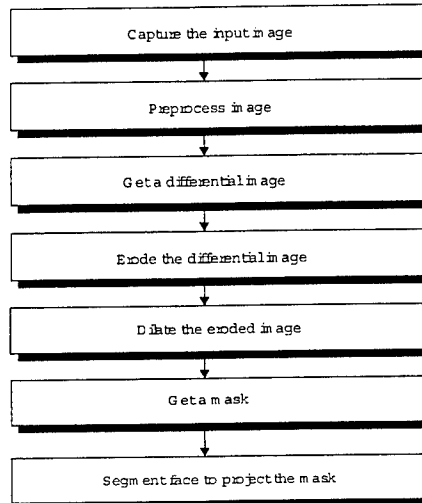


Fig. 1 Flow chart for facial image segmentation

Further since the size of human faces in regular distance is similar, square area should be defined to include eyebrow, eyes, nose and cheek like Fig. 2 and it is considered as the feature area of face recognition. After the conversion of discrete wavelet of extracted feature areas, extract the features.  $a(a=c+e+f)$  in Fig. 3 indicates horizontal distance of characteristic area and  $b$  the vertical distance.

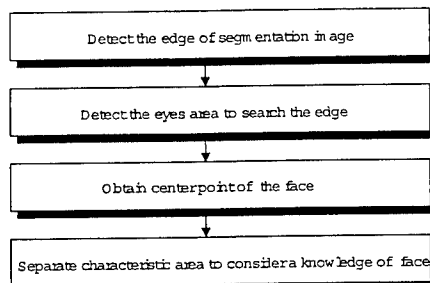


Fig. 2 Flow chart for characteristic detection

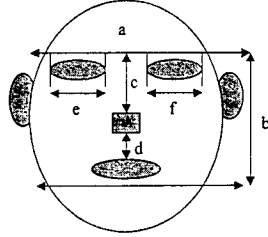


Fig. 3. The normalized distance for characteristic detection.

### 3. FACE RECOGNITION ALGORITHM BASED ON WAVELET TRANSFORM

This chapter proposes the method of recognizing the face image by using neural network after extracting the characteristics of face in wavelet conversion area. Fig. 4 shows the algorithm flow chart of face recognition. Wavelet is obtained by dilate/translating mother wavelet  $\psi(x)$  and it is shown in Expression (2).

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a}} \Psi\left(\frac{x-b}{a}\right) \quad (2)$$

where,  $a$  is scaling parameter,  $b$  translation parameter and  $\frac{1}{\sqrt{a}}$  a normalization factor. Wavelet conversion  $Wf(s)$  of random signal  $f(x)$  is defined as convolution integration with original signal  $f(x)$  and wavelet  $\psi(x)$  as shown in Expression (3).

$$Wf(x) = f * \varphi(x) \quad (3)$$

In addition, when  $\psi(x,y) = \psi(x)\psi(y)$  is two-dimensional scaling function in wavelet conversion of two-dimensional function  $f(x,y)$  and one-dimensional wavelet with one-dimensional scaling function  $\psi(x)$  is  $\psi(x)$ , two-dimensional wavelets can be explained to separable multi-resolution approximation as shown in Expression (4).

$$\begin{aligned} \varphi^A(x,y) &= \phi(x)\phi(y) \\ \varphi^V(x,y) &= \phi(x)\phi(y) \\ \varphi^H(x,y) &= \phi(x)\phi(y) \\ \varphi^D(x,y) &= \phi(x)\phi(y) \end{aligned} \quad (4)$$

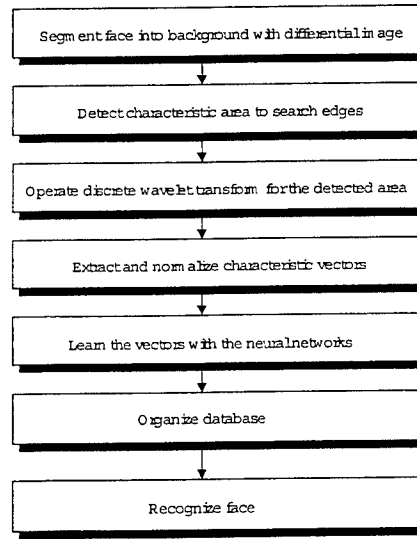


Fig. 4 Flow chart for face recognition algorithm

Two-dimensional signal divided into normal orthogonal distance of Expression (4) is divided into frequency component of spatial direction. The extraction process of feature parameters using wavelet conversion divides two input image signals with resolution of  $[256 \times 256 \times 28]$  into characteristic areas with pixel of  $91 \times 91$  using differential image and edge feature. Then DWT coefficient matrix is obtained with DWT. The following Fig. 5 shows the distribution of coefficient matrix of 4 level DWT, where cA4 means coefficient matrix of 4 level low frequency, cH(i) horizontal high frequency coefficient matrix of (i) level, cV(i) vertical high frequency coefficient matrix of (i) level and cD(i) diagonal high frequency coefficient matrix of (i) level. Feature vectors are extracted by using such coefficient matrixes. When four level DWT of segmented facial characteristic areas is performed, cA4, cH4, cV4 and cD4 which are coefficient matrixes with the size of  $6 \times 6$  are obtained, where cA4 is four level low frequency coefficient matrix, cH4 horizontal high frequency coefficient matrix, cV4 vertical high frequency coefficient matrix and cD4 diagonal high frequency coefficient matrix, and after analyzing the distribution features of these matrixes, extract feature vectors. First to examine the distribution of coefficient matrix of the same person's learning image, four level DWT is performed with four sample learning images. Then get the absolute values of coefficients of four coefficient matrixes with the size of  $6 \times 6$ , extract 36 feature vectors and normalize them between +1 and -1. Finally after calculating mean square error of normalization vector extracted from cA4, cH4, cV4 and cD4, utilize it as learning vector of neural network based on the size of error.

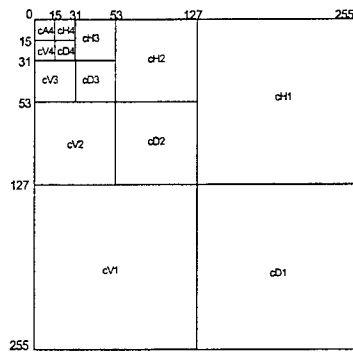


Fig. 5 Four Level DWT Coefficient Matrix

## 4. SIMULATION

### 4.1 Face Segmentation and characteristic Detection

This study acquires the input image in the same distance from CCD camera at regular illumination condition with gray scale 256 level of  $256 \times 256$  and simulation is performed on PC(266MHz). For this, the noise of input image like Fig. 6 and 7 is rejected by using gaussian filter, differential image is obtained and then the face is segmented in background image.



Fig. 6 Input Image(I)



Fig. 7 Input Image (II)

Then the differential image of two input images is obtained. Fig. 8 shows the resulting differential image and Fig. 9 is binary image.



Fig. 8 Differential Image



Fig. 9 Binary Image

Since the differential image cannot be immediately used as mask, it is binary as shown in Fig. 9 and the boundary is reduced. But since the reduction of boundary results in that of face area, the eroded(reduced) image is dilated as shown in Fig. 10. Fig. 11 is mask image generated by examining the pixel value from dilated image.



Fig. 10 Dilated Image



Fig. 11 Mask Image

And mask image is projected on original image with face and the face is segmented in the background as shown in Fig. 12.



Fig. 12 Segmentation Image



Fig. 13 Edge Image

After detecting edge using sobel operator in segmented face image, detect the position of eyes and eyebrow by horizontal distribution of edge components and segment the lower part of eyebrow. Then get the vertical distribution of edge and the center line of face. Finally, the characteristic areas are detected on the basis of the knowledge of human face as shown in Fig. 14.

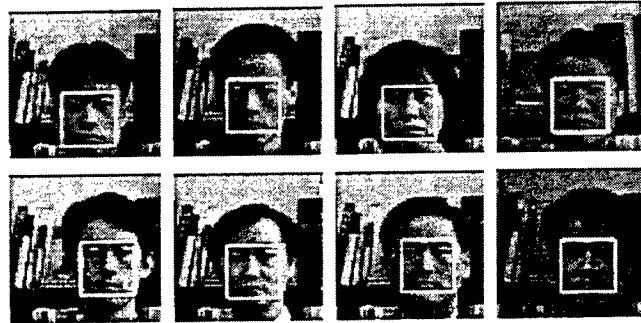


Fig. 14 Facial Characteristic Area

#### 4.2 Face Recognition Experiment

Fig. 15(a) ~ 15(d) are characteristic area separated from four experimental images of the same figure and their size is  $91 \times 91$  respectively.

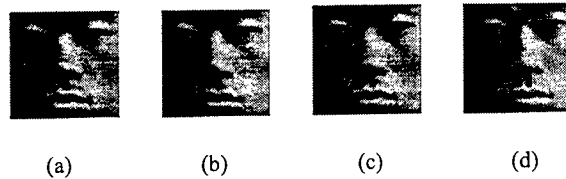


Fig. 15 Example of Separated Characteristic Image

Fig. 16 shows the image of coefficient matrixes of four level wavelet conversion cA4, cH4, cV4 and cD4 and the

information of original image is concentrated on these four coefficient matrixes. The size of these coefficient matrixes is  $6 \times 6$  respectively. For feature extraction, four learning samples of the image in Fig. 15(a)~15(d) are selected, they are converted into four level wavelet and then cA4, cH4, cV4 and cD4 are obtained. Then after analyzing normalization data distribution of the same coefficient matrixes, extract characteristic vector and it is the input vector of neural network by normalizing it between +1 and -1. Fig. 17 shows the distribution condition of normalized characteristic vector. Normalization is performed after taking the absolute value of  $6 \times 6$  coefficient matrix obtained from four sample images, where horizontal direction is the number of normalized characteristic vectors (36) and vertical direction is the range of normalization between +1 and -1. Error among normalized vectors of the same coefficient matrix is obtained to select input vector of neural network of coefficient matrixes from four sample images like figure 15(a)~15(d). First, RMSE of normalized characteristic vector is obtained in cA4. Then RMSE of cH4, cV4 and cD4 is obtained through the same process. The following Table 1 shows RMSE of each coefficients.

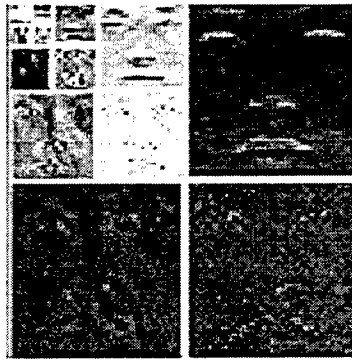


Fig. 16 Four Step Decomposition Image

Total 108 of each 36 from cA4, cH4 and cV4 are extracted as characteristic vectors based on Table 1 and these are the input vectors of neural network. Fig. 18 shows the distribution of 108 neural network input vector extracted from four sample images of the same person.

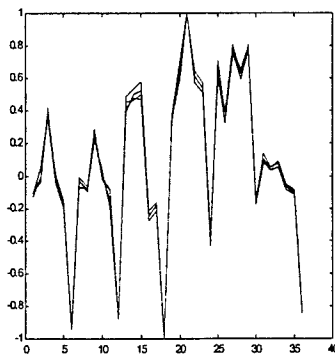


Fig. 17 Normalized Characteristic Vectors

Table 1. RMSE of normalized feature vectors

Class	cA4	cH4	cV4	cD4
RMSE	0.0035	0.00486	0.00638	0.00658
	75	7	1	0

To confirm the validity of characteristic vector extraction, four images of three persons are obtained, wavelet conversion is conducted and the normalized results of coefficient matrix are shown. For this, four sample images of experimental image like Fig. 19(a), 19(b) and 19(c) are selected, four level wavelet conversion of these images is performed,  $6 \times 6$  coefficient matrix is normalized and 36 characteristic vectors are extracted. And RMSE among the same coefficients is obtained.

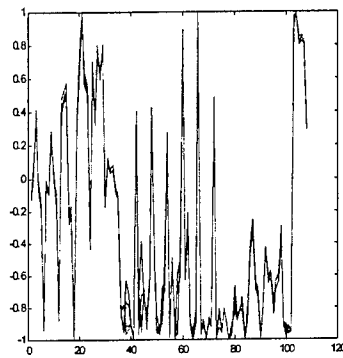


Fig. 18. Extracted Input Vectors

Table 2, 3 and 4 are RMSE of normalized vectors. The largest error is found in normalized vector of cD4.

Table 2. RMSE of normalized feature vectors

Class	cA4	CH4	cV4	cD4
RMSE	0.0030	0.00556	0.00304	0.00672
	76	5	2	6

Table 3. RMSE of normalized feature vectors

Class	cA4	CH4	cV4	cD4
RMSE	0.0017	0.00383	0.00219	0.00622
	88	0	6	4

Table 4. RMSE of normalized feature vectors

class	cA4	cH4	cV4	CD4
RMSE	0.0008	0.00292	0.00108	0.00347
	48	8	2	6



Fig. 20(a), 20(b) and 20(c) shows the neural network input vector extracted from Fig. 19(a), 19(b) and 19(c).

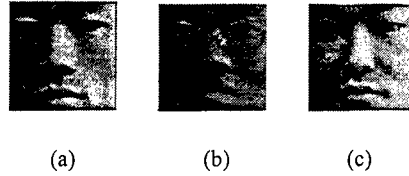


Fig. 19 Example of experimental images.

Such 108 normalized input vectors are multi-layer neural network. Learning algorithm of neural network uses error Backpropagation learning algorithm. The error of output layer is 0.005 and learning rate is 0.7. Weight of network generated after learning and feature vector of input comparative image are operated and the recognition is performed by comparing the error of output layer. Then when the error of output layer is less than 0.005, it is judged as the same person.

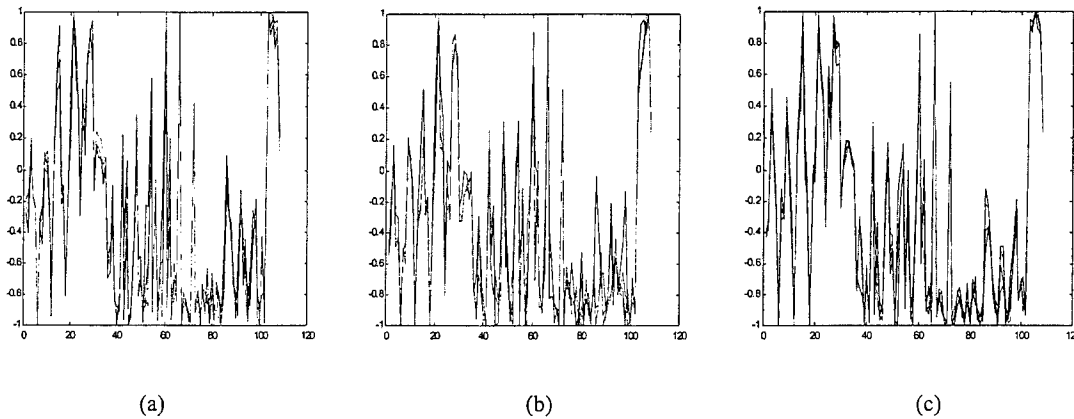


Fig. 20. Extracted input vectors

#### 4.3 Comparison and Examination

The simulation results of proposed algorithms are as follows.

- 1) Facial image can be segmented 100% in proposed experimental environment by using face segmentation algorithm based on differential image.
- 2) More elaborate feature vector can be obtained than by using DCT and the recognition rate increases 4%. Fig. 21 shows that the normalization characteristic of feature vectors extracted after wavelet conversion is more evenly distributed than the distribution characteristic of extracted after performing DCT of the same person.

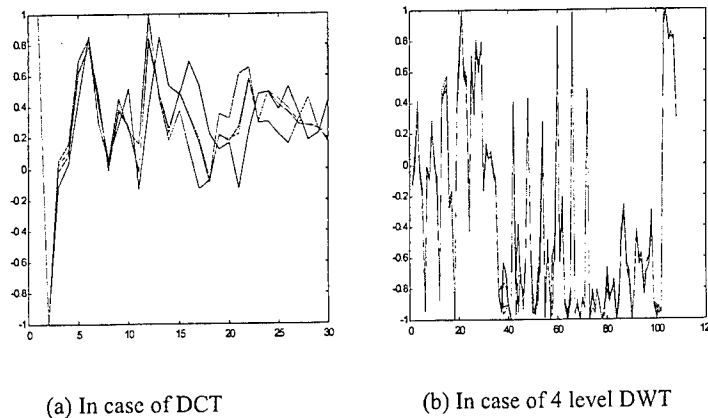


Fig. 21 Distribution of Characteristic Vectors

- 3) Features are extracted by converting the data of spatial area into frequency area and data quantity can be reduced. Also data quantity for operation is reduced by handling characteristic area not the whole face in performing DWT.
- 4) Since the wrong recognition of included face image occurs, the compensation of inclination is required. And algorithm such as pass word input or fingerprint recognition is required to apply it to perfect complementary system.

## 5. CONCLUSIONS

We have presented our experiments about face segmentation method based on differential and face recognition using discrete wavelet transform. This algorithm was proposed by following processes. First, two gray-level images was captured with 256 level of the size of  $256 \times 256$  in constant illumination. We used a Gaussian filter to remove noise of input image and got a differential image between background and input image. Second, a mask was made from erosion and dilation process after binary of the differential image. Third, facial image was divided by projecting the mask into input image. Most characteristic information of human face was in eyebrow, eyes, nose and mouth. In the reason, the facial characteristic were detected after selecting local area including this area. Forth, detecting the characteristic of segmented face image, edge was detected with Sobel operator. Then, eye area and the center of face were searched by using horizontal and vertical components of edge. Characteristic area consists of eyes, a nose, a mouth, eye brows and cheeks, was detected by searching the edge of the image. Finally, characteristic vectors were extracted from performing DWT of this characteristic area and are normalized it between +1 and -1. Normalized vectors was used with input vector of neural network. Through simulation results, we shown recognition rate of 100 % about learned image and 92% about test image.

## REFERENCES

- 1 T. Sakai, M. Nagao, and Fukibayashi : Line extraction and pattern recognition in a photograph : *Pattern Recognition*, Vol. 1, pp. 233 ~248, 1969.
- 2 M.D. Kelly : Visual identification of people by computer : Tech. Rep. AI-130, Stanford AI Proj., Stanford, CA, 1970.

- 3 I. Craw, H. Ellis, and J. Lishman : Automatic extraction of face feature : *Pattern Recognition Lett.*, Vol. 5, pp. 183 ~ 187, 1987.
- 4 V. Govindaraju, S. N. Srihari, and D. B. Sher : A computational model for face location : *Proc. 3rd Int. Conf. on Computer Vision*, pp. 718 ~ 721, 1990.
- 5 S. A. Sirohey : Human face segmentation and identification : Tech. R. CAR-TR-695 : Center for Automation Research, Univ. of Maryland, 1993.
- 6 Myung-Kil Lee, Joo-Sin Lee, "A Study on Pattern Recognition using DCT and Neural Network" *The Journal of the Korean Institute of Communication Science*, Vol. 22 No. 3, pp. 481-492, 1997.
- 7 Myung-Kil Lee, Joo-Sin Lee, "Pattern Recognition of SMD IC using Wavelet Transform and Neural Network" *Journal of the Korea Institute of Telematics and Electronics*, Vol 34-B No 7, pp. 768-777. 1997.

# Some contributions to wavelet based image coding

Yi-Wei Li, Kuo-Shu Chang, Loon Shan Yan and Day-Fann Shen

Department of Electrical Engineering  
Yunlin University of Science and Technology  
Taiwan

## ABSTRACT

Four key issues in wavelet zero-tree based image coding are investigated and presented, they are (1) Fast wavelet transform that save 1/2 and 3/4 processing for one dimensional signal and two dimensional signals respectively. (2) The selection of the best wavelet filters that yields best performance (PSNR vs. Bit rate) for most common seen images. (3) Recommendation of number of wavelet scales (or frequencies) for image coding by experiments and analysis.

Keywords: wavelet transform, image coding, wavelet scales, filters.

## 1. INTRODUCTION

Wavelet zero-tree based approach has been adopted in JPEG2000 image coding standard [1]. In this paper, We discuss four important issues in a wavelet zero-tree based image coding and present our solutions for improvements. Firstly, Wavelet decomposition/synthesis are the basic process in any wavelet based image coding, it is highly desired to speed up the process without sacrifice the precision. By integrating the convolution and the sub-sampling process, our approach is able to speed up the wavelet transform up to 1/2 for one dimensional signal and 3/4 for two dimensional signal. The correctness of the results is verified using MATLAB wavelet toolbox.

Secondly, it is known that the main reason that discrete wavelet transform (DWT) outperforms discrete cosine transform (DCT) in image coding is that wavelets of finite duration forms a set of better basis than the periodic cosin basis in representing the signals and the images. Many wavelet filters have been proposed [2] [3], it is interested to know which wavelet filters are the best for image coding. We have tested all 77 sets of wavelet coefficients provided in MATLAB wavelet toolbox (version 1) for their performances (Bit-rates vs. PSNRs) using a zero-tree based coding technique [4]. The results indicate that two set of wavelet filters Bior-4.4 and Bior-5.5 are consistently the best for three test images of different complexity: Lena, Goldhill and Pepper. Both using technique [4]. The results indicate that two sets of wavelet filters are bi-orthogonal and linear phased filters. It is known that Bior-4.4 filter is also recommended for the coding of finger print by CIA [5].

Finally, one wavelet decomposition generate four sub-bands (LL, HL, LH, HH) of the same scale, the octave wavelet subbands of different scales can be obtained by decomposing the lowermost subband (LL) repeatedly, an wavelet subbands of three scales are shown in Figure 1. One key parameter in wavelet transform is the optimal number of scales for image coding. We investigate this issue via performance evaluation with a brief analysis. We recommend that for image of higher complexity, like Goldhill, 3 scales are the best while for images of medium and low complexity (Lena and Pepper), 4 scales of decomposition is the best choice.

## 2. FAST WAVELET ANALYSIS/SYNTHESIS

Signal analysis and synthesis based on the wavelet can be efficiently implemented using a pair of QMF (Quadratic Mirror Filters) filters as proposed by Mallat[2]. A one dimensional one scale (level) wavelet analysis and synthesis process is shown in Figure 2. Wavelet subbands with octave scales can be obtained by successively decomposing the lowest frequency subband. Since the bandwidth of the low-passed and high-passed signal is halved, they can be down-sampled by a factor of 2. Let  $m$  be the length of the input signal  $X(n)$  and  $n$  be the length of an analysis filter,  $m$  is usually an even number while  $n$  can be either even or odd. The length of the signal immediately after circular convolution ( $X(n)$  is regarded as an periodic signal) with the analysis filter and before down-sampling is  $m+n-1$ . It is normally required that length of  $Y_l(n)$  and  $Y_h(n)$  be  $m/2$ , therefore, the filter output must be truncated to

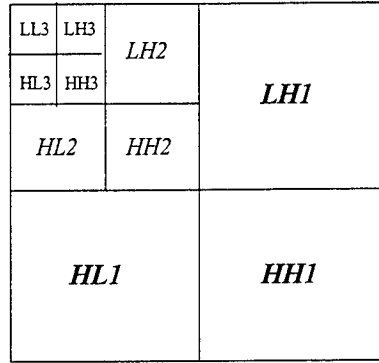


Figure 1. 3-scale wavelet decomposition

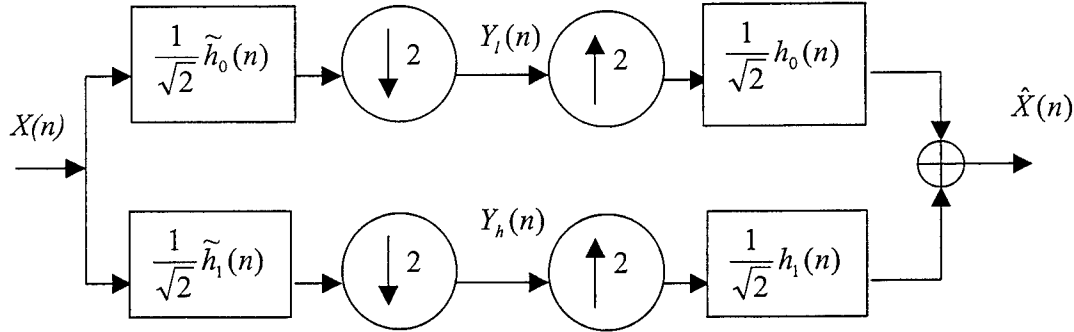


Figure 2. One dimensional one scale (level) wavelet analysis and synthesis process using QMF.

$m$  before down-sampling. A common practice to achieve this requirement is by removing  $(n-1)/2$  samples at both end of the filtered output if  $n$  is even or remove  $n/2$  samples at one end and  $n/2-1$  at the other end. The length truncated signal is then down-sampled to obtain  $Y_l(n)$  and  $Y_h(n)$ . The function “dwt2per” in MATLAB wavelet toolbox is designed to perform the truncation.

We proposed a method to improve the process of generating  $Y_l(n)$  and  $Y_h(n)$ , the idea is that convolution and down-sampling can be done in one process. Figure 3 is an example used for illustration, here  $n=4$  is used. Note that if the length of the analysis and synthesis filters are not equal, zeros be added to the shorter filter.  $X(n) = [X(1) \dots X(m)]$  is treated as a periodic signal of period  $m$ ,  $f(n) = [f(1) \dots f(4)]$  is the reversed ordered filter coefficients for the convolution process.

The low-pass or high-pass filter output  $Y(k) = \sum_{j=1}^{j=n} f(j)x(2(k-1)+j)$  for  $k=1 \dots m/2$ . Note that only  $m/2$  points

of outputs are calculated. This simple idea reduces the computations to  $(1/2)^k$  of the original method.

Moreover, the signal truncating problem is now eliminated. The reconstruction process includes up-sampling (inserting zeros between signal data) to form a periodic signal  $Y'(n)$  of period  $m$ . To reconstruct the signal

$X'(k)$ ,  $X'(k) = \sum_{j=1}^{Round[n/2]} f(2j-1)Y'(k+2j-2)$  if  $k$  is odd and  $X'(k) = \sum_{j=1}^{Round[n/2]} f(2j-1)Y'(k+2j-1)$ , if

$k$  is even. The example with filter length  $n=4$  is shown in Figure 4, where  $f(1)$ - $f(4)$  are the reverse ordered reconstruction filter coefficients. The output signals from low-pass and high-pass reconstruction filters are then summed up to reconstruct the original signal.

[illegible]

Data <sub>i</sub>	X(1)	0	Y(2)	0	Y(3)	0	Y(4)	0	K K K K K R	Y(m/2)	0	Y(1)	0	Y(2)
j	i J	i J	i J	i J	i J	i J	i J	i J	K K K K K Ki J		i J			
filter <sub>i</sub>	f(1)	f(2)	f(3)	f(4)										
	i J	i J	i J	i J							i J	i J		
shift 1		f(1)	f(2)	f(3)	f(4)									
points	i J	i J	i J	i J	i J						i J	i J		
			f(1)	f(2)	f(3)	f(4)								
	i J	i J	i J	i J	i J	i J					i J	i J		
:	:		:		:						:	:		
									f(1)	f(2)	f(3)	f(4)		
										f(1)	f(2)	f(3)	f(4)	
											f(1)	f(2)	f(3)	f(4)
	X(1)	X(2)	X(3)		K K K K K K				X(m-2)	X(m-1)	X(m)			

### 3. SELECTION OF WAVELET FILTERS FOR ZERO-TREE BASED IMAGE CODING

Separable wavelet filter and separable down-sampling allow very efficient implementation of wavelet transform, since the 2-D filtering can be break down to two cascaded 1-D filtering process, the total computations is reduced from  $O(N^4)$  to  $O(N^3)$ . However, the drawback of separable wavelet transform is that only rectangle pieces of the spectrum can be isolated, this is because separable 2-D filtering is the product of 2 1-D filters. Different shapes of spectrum other than rectangle would require non-separable 2-D filters, which allow better coding performance at the cost of higher computational complexities and may have the stability problem. [6]. In practice, most wavelet transforms are using separable filters, all the 77 sets of wavelet filters in MATLAB wavelet toolbox are all separable filters.

Orthogonal wavelet filters implement “unitary transform” between input and the decomposed sub-bands. It implies that energy, distortion as well as bit rate are conserved between the input signal and the decomposed sub-band signals. The bit allocation algorithm can be implemented easily using this property. However, it has proven that linear phase and orthogonality are mutual exclusive in a separable FIR system. Linear phase using orthogonal filter can be achieved under the following conditions. (1) An orthogonal filter of sufficient length can be made almost linear phase. (2) Non-separable orthogonal filters may have linear phase and (3) Orthogonal IIR filter allow linear phase. In practice, most implementation of wavelet transforms adopt well designed Bi-orthogonal filter for linear phase while keep the conservation property as close to an orthogonal filter as possible.

An orthogonal filter with a certain number of zeros at the aliasing frequencies ( $\pi$  in two channels case) is called regular if its iteration tends to a continuous function. A filter of high regularity improves coding gain the compression artifacts is less objectional. Filters of low regularity cause poor coding performance, moderate regularity improves the performance significantly, however, higher regularity can only improve the performance a little. For bi-orthogonal filtering, only either the analysis or the synthesis filter can be regular. To minimize the visibility of objectional artifacts, it is preferred to have a regular synthesis filter.

Wavelet filter of higher regularity requires longer filters, however, longer filters have the following drawbacks, (1) It requires more computations. (2) It tends to spread coding error around (3) Longer filter tends to have more zero-crossings which causes more ringing artifacts around edges. For image compression, shorter and smoother filters are preferred [6].

To select the best wavelet filter available for zero-tree based image coding, we conduct an experiment to evaluate the performance for all 77 sets of wavelet filters collected in MATLAB wavelet toolbox version 1. The 256x256 Lena image and JZW (JND based zero-tree wavelet) as proposed by Shen and Yan in 1998 [4] are used in the initial performance evaluation process. Among the best 9 sets of wavelet filters (in PSNR), 6 sets are biorthogonal filters (Bior2.2, bior2.4, Bior2.6, bior4.4, bior5.5, bior6.8) and 3 sets are symlet (sym4, sym5, sym6). The PSNR vs. bit rate curves for these 9 set filters are shown in Figure 5.

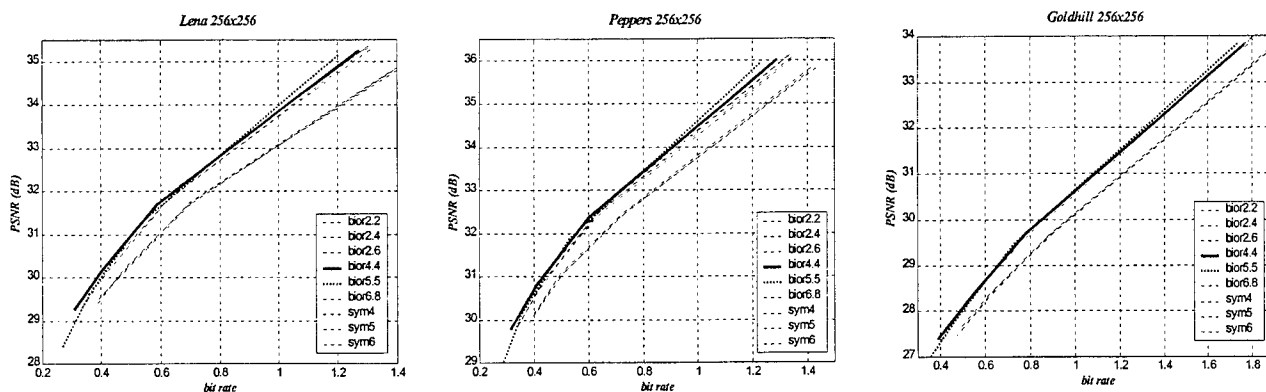


Figure 5. Performance (PSNR vs. Bit-rate) of the best 9 wavelet filters in MATLAB wavelet toolbox

Among the 9 sets of the best wavelet filters, bior4.4 and bior5.5 consistently outperform the others for Lena, Pepper and Goldhill images in the range of PSNR 28 to 35. By further examining the detailed performance data for bior4.4 and bior5.5 in Table I, we found that bior4.4 performs better in the range of low to medium bit rate (below 0.8 bpp) while 5.5 performance better for higher bit rate (0.8 bpp and above). The filter coefficients for bior4.4 (9/7 filter) and bior5.5 (11/9 filter) are listed in Table II and Table III.

Table I. Bit rate vs. PSNR data for Bior4.4 and Bior5.5 filters.

Complexity Bit rate factor PSNR		Goldhill (high)		Lena (median)		Peppers (Low)	
		bior4.4	bior5.5	bior4.4	bior5.5	bior4.4	Bior5.5
$\frac{1}{8}$ $\frac{1}{8}$ 1	bpp	1.7635	1.7341	1.2725	1.2046	1.2898	1.2270
	PSNR (dB)	33.8227	33.8407	35.2504	35.143	36.0072	35.9261
$\frac{1}{8}$ $\frac{1}{8}$ 3	bpp	0.7805	0.7253	0.5864	0.5365	0.6025	0.5296
	PSNR (dB)	29.7120	29.4590	31.7021	31.2906	32.3795	31.8880
$\frac{1}{8}$ $\frac{1}{8}$ 5	bpp	0.5274	0.4650	0.4016	0.3548	0.4096	0.3568
	PSNR (dB)	28.2795	27.8079	30.1586	29.5543	30.8100	30.1360
$\frac{1}{8}$ $\frac{1}{8}$ 7	bpp	0.3881	0.3331	0.3092	0.2659	0.3147	0.2737
	PSNR (dB)	27.3825	26.8423	29.2472	28.3659	29.7608	28.8183

Table II. Filter coefficients for Bior4.4 (9/7 taps QMF)

n		0	$\pm 1$	$\pm 2$	$\pm 3$	$\pm 4$
Decomposition Filter	$\tilde{h}_0(n)$	-0.557543	0.295636	0.02877	-0.045636	0
	$\tilde{h}_1(n)$	0.602949	0.266864	-0.07823	-0.016864	0.026749
Reconstruction Filter	$h_0(n)$	-0.602949	0.266864	0.78223	-0.016864	-0.026749
	$h_1(n)$	0.557543	0.295636	-0.02877	-0.045636	0

Table III. Filter coefficients for Bior5.5 (11/9 taps QMF)

n		0	$\pm 1$	$\pm 2$	$\pm 3$	$\pm 4$	$\pm 5$
Decomposition Filter	$\tilde{h}_0(n)$	0.6360	-0.3372	-0.0661	0.0967	-0.0019	-0.0095
	$\tilde{h}_1(n)$	0.5209	0.2444	-0.0385	0.0056	0.0281	0
Reconstruction Filter	$h_0(n)$	0.5209	-0.2444	-0.0381	-0.0056	0.0281	0
	$h_1(n)$	0.6360	0.3372	-0.0661	-0.0967	-0.0019	0.0095

The selection of Bior4.4 and Bior5.5 filters are consistent with the above requirements analysis of an idea wavelet filters in image coding. Both are separable, bi-orthogonal, linear phase, short and smooth with moderate regularity.

#### 4. NUMBER OF WAVELET SCALES IN WAVELET ZERO-TREE BASED IMAGE CODING

Theoretically, octave subbands can be obtained through repeated wavelet analysis on the low-frequency subband until a single pixel is reached [6], thus a maximum number of scales for a  $M \times M$  image is  $N_{Max\_scales} = \log_2 M$ . It is nature to ask the question: How many wavelet scales is most suitable for image coding? There are few papers discuss this problem, SPIHT [7] arbitrarily chooses 5 scales, while Rajala et al. [8] choose 3 scales. In this paper, we investigate the relation between number of wavelet scales and coding performance in image coding. A zero-tree based coding technique JZW [4] is adopted for the performance evaluation purpose. The wavelet transformed coefficients in each sub-band are quantized by a JND weighted step size  $\Delta_{subband}$ , this process is called JND\_SQ (or JND based Scalar Quantization).  $\Delta_{subband}$  for each sub-band are derived from extensive experiments and has larger value for higher frequency sub-band and has smaller value for lower frequency sub-band. Thus, wavelet coefficients in higher frequency (smaller scale) subbands are quantized more coarsely while the lower



frequency (larger scale) subbands are quantized relatively finer. After JND\_SQ, wavelet coefficients smaller than  $1/2$  of the step size are quantized to zeros. As a result, many zeros are produced in each sub-band, especially at higher frequency sub-bands where step size is large and coefficients values are smaller. It is known that the more zeros, the higher coding efficiency in zero-tree based image coding. In addition, it is important to know that the JND step sizes are carefully derived that the reconstructed image after JND\_SQ maintaining a visually loss-less quality even under viewing condition in dark room and at any viewing distance. Although JZW is designed to optimize the perceptual image quality, several zero-tree based image compression techniques are proposed to enhance the coding performance, as a result, JZW outperforms SPIHT even in terms of MSE or PSNR. However, it is noted that JZW does not implement the embedded property as in EZW and SPIHT and that the embedded property can be achieved by passing the JND quantized wavelet coefficients to EZW or SPIHT.

It is noted that the lowest frequency (the coarsest scale) band (LL band) is the most important subband to human perception. Also, most coefficients in LL band have very large values and unlikely to be zeros, it is not efficient to include LL band in the zero-tree scanning. For these two reasons, JZW encodes the LL band separately using loss-less DPCM. Wavelet coefficients in other higher frequency subbands can be efficiently encoded using our zero-tree encoding scheme which is derived from EZW and [improved version of EZW by SAID and PERALMAN, as indicated by song in 4/28/2000 presentation]. In JZW, each zero-tree is encoded by list of coefficient states (LCS) and list of coefficient values (LCV). To simplify the implementation, JZW omits the sophisticated embedded property, i.e. JZW encodes the full value (JND quantized) of the coefficients in one pass rather than the bit plans in multiple passes. JZW also simplifies the 4 possible states to 3 possible states for zero-trees. The three states are {ST (Significant Tree), SR (Significant Root) and ZTR (ZeroTree Root)}. A ST coefficient has at least one non-zero descendent; a SR coefficient is non-zero itself but all descendents are zeros; A ZTR coefficient and all its descendents are zeros. The children of a ST coefficient may have three possible states {ST, SR and ZTR}; Children of {SR and ZTR} must be ZTRs and their descendents must be all zeros and can be skipped in the coding process. It is noted that the more zero coefficients the higher coding efficiency and that JND\_SQ can effectively reduce those coefficients smaller than  $1/2 \Delta_{subband}$  into zeros without degrading the perceptual quality.

Since we are interested in determining the wavelet scales, the wavelet decomposition of 3, 4 and 5 scales (corresponding to total of 10, 13, 16 wavelet subbands) are generated for a set of 6 test images (Goldhill, Lena, Pepper of 512x512 and 256x256). The performance (Bit rate vs. image quality) of JZW on each test image of different scales are recorded. Table IV lists the best wavelet scales for the 6 test images.

Table IV. Best wavelet scales for the 6 test images

complexity size \ Scale	Goldhill (high)	Lena (moderate)	Peppers (Low)
512x512	3	4	4
256x256	3	3	3

Based on the data in Table IV, we conclude that 3 or 4 scales are most suitable for common seen images. For images of lower resolution (256x256) and images of higher complexity (Goldhill), 3 scales (10 subbands) is the best; For images of higher resolution (512x512) and moderate or lower complexity, 4 scales (13 subbands) is the best.

We further investigate the reasons behind the optimal decomposition levels. Consider a 5-levels wavelet zerotree shown in

Table V. If a node at level 5 (the lowest frequency band) is an SR(condition 1), then 4-level decompositions (4 scales) for the same image would require one more symbol than 5 level decomposition. However, if a node at level 5 is ZTR or SR, then 4-level decomposition (4 scales) can save at least three symbols than 5-level decomposition (5 scales). There are tradeoffs in terms of required number of symbols. How to choose the optimal levels of wavelet decomposition depends on the percentage of condition 1 and condition 2 nodes.

After N level decompositions, if the percentage of condition 2 nodes (ZTR or SR) and condition 1 (ST nodes) are 25% and 75% respectively, then N level (N scales) and N-1 level decomposition requires about the same number of symbols. Therefore, the rule of thumb is that *if the percentage of condition 2 (ZTR or SR) is less than 25% after N levels decomposition, then the optimal levels of decomposition is N-1*. To illustrate, we take 512x512 and 256x256 "Lena" as an example. Table VII shows the percentage of condition 2 for different levels of wavelet decomposition. We may find the optimal levels of wavelet decomposition for 512x512 as well as 256x256 Lena using the above rule of thumb. ST node often appears on the edges while SR/ZTR appears in smooth areas of an image. The more complex an image is, the more edges it has. So there is fewer percentage of ZTR /SR in a complex image, therefore, fewer wavelet decomposition levels are necessary.

In general, most of the commonly seen natural images have low or medium complexity. Therefore, 3-levels wavelet decomposition is recommended for 256x256 or the similar sizes such as CIF (352x288) or QCIF (176x144), while 4-levels wavelet decomposition is recommended for 512x512 images or the similar sizes respectively.

Table V. Example of required symbols

		condition 1					condition 2				
		ST					ZTR or SR				
Level 5		1					1				
Level 4		2	3	4	5		2	3	4	5	
		ST	ST	ZTR	SR		ZTR	ZTR	ZTR	ZTR	
	node:	1	2	3	4	5	1	2	3	4	5
(5-scale)	LCS:	1	1	1	0	2	0 or 2	x	x	x	x
	LCV:	v	v	v	x	v	x or v	x	x	x	x
	Total symbols:	9					1 or 2				
(4-scale)	LCS:	x	1	1	0	2	x	0	0	0	0
	LCV:	Dc	v	v	x	v	Dc	x	x	x	x
	Total symbols:	8					5				

Dc: coefficient in the lowest subband  
v: significant value  
X: don't code

Table VI. Percentage of condition 2

Lena		
	512x512	256x256
Level 1	99.83 %	99.26%
Level 2	92.55%	85.74%
Level 3	69.45%	52.96%
Level 4	36.16%	18.16%
Level 5	8.07%	1.56%

## 5. CONCLUSION

We have presented our contributions on four key issues in wavelet zero-tree based image coding. They are (1) Fast wavelet transform that save 1/2 and 3/4 processing for one dimensional signal and two dimensional signals respectively. (2) The selection of the best wavelet filters that yields best performance (PSNR vs. Bit rate) for most common seen images. (3) Recommendation of number of wavelet scales for image coding by experiments and analysis

## 6. REFERENCES

- [1] [www.JPEG.org/JPEG2000.htm](http://www.JPEG.org/JPEG2000.htm)
- [2] Stephane G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no.7, pp.674-693, July 1989.
- [3] Ingrid Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Tran. Information*

*Theory*, vol. 36, no. 5, Sep. 1990.

- [4] Day-Fann Shen and Loon-shan Yan, "JND measurements and wavelet based image coding", *SPIE International Optoelectronics Exposition*, July 1998
- [5] A. B. Watson and G. Y. Yang, "Visibility of Wavelet Quantization Noise," *IEEE Trans.* Vol. 6, no. 8, August 1997
- [6] Martin Vetterli and Jelena Kovacevic, "Wavelets and Subband Coding"
- [7] Amir Said and William, "A new, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees". *IEEE Transactions on circuits and systems for video technology*, vol. 6, pp. 243-250, June 1996.
- [8] Robert E. Van Dyck and Sarah A. Rajala, "Subband/VQ Coding of Color Images With Perceptually Optimal Bit Allocation". *IEEE Transactions on circuits and systems for video technology*, vol. 4, pp. 68-82, Feb. 1994.

# Fast ITTBC using Pattern Code on the Subband Segmentation

Sung Shick Koh<sup>a</sup>, Han Chil Kim<sup>a</sup>, Koo Young Lee<sup>a</sup>, Hong Bin Kim<sup>a</sup>  
Hun Jeong<sup>b</sup>, Gang Seok Cho<sup>b</sup>, Chung Hwa Kim<sup>\*a</sup>

<sup>a</sup>Dept. of Electronic Engineering, Chosun University, Korea

<sup>b</sup>Dept. of Electronic Chosun College of Science & Technology, Korea

## ABSTRACT

Iterated Transformation Theory-Based Coding (ITTBC) suffers from very high computational complexity in encoding phase. This is due to its exhaustive search. In this paper, our proposed image coding algorithm preprocess an original image to subband segmentation image by wavelet transform before image coding to reduce encoding complexity. A similar block is searched by using the 24 block pattern codes which are coded by the edge information in the image block on the domain pool of the subband segmentation.

As a result, Numerical data shows that the encoding time of the proposed coding method can be reduced to 98.82% of that of Joaquin's method, while the loss in quality relative to the Jacquin's is about 0.28dB in PSNR, which is visually negligible.

**Keywords:** ITTBC, wavelet transform, subband, pattern code, domain, range, quadtree

## 1. INTRODUCTION

Digital image coding techniques are very important in various areas for efficient storage or transmission of images. It is noticed that the encoding method of image data by fractal and subband using the wavelet theory as the efficient encoding technology with preserving the image loss and brightness.

To contract image data, so far suggested representative coding methods are transform coding, vector quantization, fractal coding and wavelet transform coding<sup>1</sup> that is being studied recently. A fractal scheme introduced by Mandelbrot<sup>2</sup> at first and applied gradually to image coding by M. Barnsley<sup>3</sup> and applied by using Iterated Function System (IFS) by Jacquin<sup>4</sup> and Fisher<sup>5</sup>, which include a self-similarity in the given image. ITTBC<sup>6,7,8</sup> method uses that all images have large or small self-similarity in those images. The essence of the compression process is the pairing of each range block to a domain block

---

\* Correspondence: Email: jhdkim@mail.chosun.ac.kr; Telephone: +82-62-230-7068 ; Fax: +82-62-255-6311

such that the difference between the two, under an affine transformation, is minimal. This involves a lot of searching.

ITTBC method could typically be compressed to high compression ratio while still looking pretty good. Nonetheless, this method has the problem of computation burden because searching time is required to obtain the domain block which is adequate to a range block.

In this paper, reducing the searching time for the similar block while keeping the high compression ratio and the high quality of the decoding image, our proposed image coding method at first allows of subband segmentation<sup>9</sup> of original image by wavelet transform using Daubechies's filter, second decomposes all subband images to quadtree multiresolution blocks according to image information, third codes each quadtree multiresolutioned domain block according to the pattern of the edge information in the block and forth searches adaptively similar domain block on the low frequency band which the important image information is compacted in.

## 2. SUBBAND SEGMENTATION BY WAVELET TRANSFORM

The basic concept of the wavelet transformation means expressing a certain function  $f$  by superimposing the wavelets. Wavelet transform is to base an analysis on one function  $\Psi(x)$  that would be well localized in both time and frequency. This function was then dilated and translated to form a family of analysis functions. There are normalized as follows

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a}} \Psi\left(\frac{x-b}{a}\right) \quad a, b \in R \quad (1)$$

Where  $a$  is a scale parameter by which the wavelet is contracted and expanded, and  $b$  is a translation parameter. Since  $a$  is a scale parameter, the scale of the basis wavelets can be controlled. And the basis wavelets can be translated anywhere by varying  $b$ .

Discrete wavelet transform is discovered by carefully choosing the analysis function and by taking  $a = 2^{-j}$  and  $b = 2^{-j}k (j, k \in Z)$  as discrete values for the parameters  $a$  and  $b$ , that one could obtain orthonormal bases for  $L^2(R)$  of the type

$$\Psi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \Psi(2^j x - k) \quad j, k \in Z \quad (2)$$

and that the expression

$$f = \sum_{j,k \in Z} \langle f, \Psi_{j,k} \rangle \Psi_{j,k} \quad (3)$$

for decomposing a function in these orthonormal wavelets covered in many function spaces. An image is divided into four subbands by wavelet transform. We can have a multiresolution as a result of iteration of wavelet transform.

### 3. PROPOSED METHOD

We propose the fast ITTBC algorithm that can reduce the encoding time and keep high compression ratio and faithfulness of the decoded image, which use the adaptive searching method by the block pattern code and band limitation on the subband segmentation.

#### 3.1 Adaptive Multiresolution on the Subband Domain Pool

In the field of image coding algorithms, the wavelet transform method which get high compression ratio can control the data information for each band divided by the weighted rate because input image is decomposed to multiresolution by visual sensibility. The subbands using in this paper by the wavelet transform can be shown by figure 1.

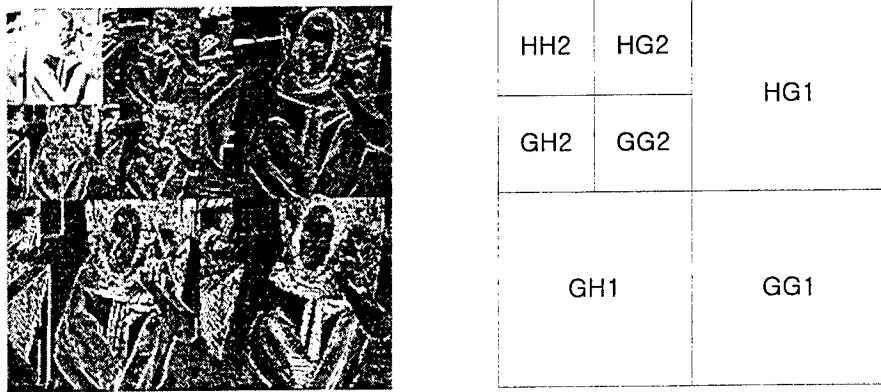


Figure 1. Subband segmentation by 2 level wavelet transform

First, we construct the decomposed subband image by 2 level wavelet transform. Second, we divide adaptively the wavelet transformed subband(HH2, HG2, GH2, HG1, GH1) into three step sub-blocks based on quadtree partitioning as shown in table 1.

The HH2 band having a large amount of image information can be divided into the first step (8 by 8) blocks. If the image information(8 by 8 blocks) exceed the setting threshold value, it means the block is dividable into the smaller blocks such as the second step (4 by 4)blocks, the third (2 by 2)blocks. In HG2 and GH2 band having a lot of vertical and horizontal image information, they can be divided into the first step (16 by 16)blocks. According to the amount of information, they can be divided into low-step blocks such as the second step (8 by 8)blocks and the third step (4 by 4)blocks. Also, the HG1 and GH1 band can be divided into the first step (32 by 32)blocks. According to the amount of information and they can be divided into the low-step blocks such as the second step (16 by 16)blocks, the third step (8 by 8)blocks. In diagonal image band GG1 and GG2 having a little amount of information, we can get the average image information.

Table 1. Design specifications of the proposed coding system

Original Image	
Name	Barbara
Resolution	256
Gray levels	256
Subband Segmentation	
Wavelet transform level	2 level WT
Filter	Daubechies's
Encoding Specifications	
Range partition type	3 step multiresolution quadtree
Range blocks	HH2 : 8 by 8, 4 by 4, 2 by 2 HG2, HH2 : 16 by 16, 8 by 8, 4 by 4 HG1, GH1 : 32 by 32, 16 by 16, 8 by 8 GG2, GG1 : Average
Domain blocks	32 by 32, 16 by 16, 8 by 8, 4 by 4

Therefore, each subband can be divided adaptively into multiresolution quadtree partition according to the amount of information. That is, most of energy is concentrated on the low resolution band. If the band has high resolution, there is small energy. By using this, we can minimize the range blocks and progress the compression ratio.

### 3.2 Block Pattern Code

After partitioning each subband until the adaptive third step multiresolution on the wavelet transformed image with edge information, we take an image information to the mean value for the first step domains.

In the second step and the third step quadtree partition domain pool, all blocks are broken up into four equal-sized sub-squares and taken to the mean values for each block. And it is lined up to a few mean values and coded for the block pattern. Shown in Figure 2.

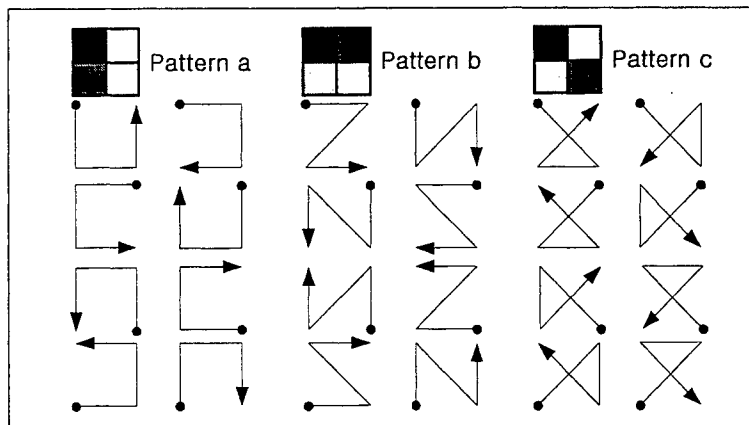


Figure 2. Block pattern code

If we code the domains by the proposed block pattern in advance according to the orientation of the edge information on

the subband segmentation, we can limit the searched domain pool into the each resolution and determine the similarity both ranges and domains on the domain pool according to the orientation index. Therefore, the range blocks are efficient to a search similar domain block because of searching for pattern code.

#### 4. SIMULATION AND RESULTS

We estimated the fast ITTBC algorithm using pattern code on the subband segmentation compared with the result of the ITTBC method using the geometric transform proposed by Jacquin's <sup>4</sup>. An original digital image "barbara"(gray level : 8 b/pixel, size : 256 by 256) shown in figure 3 was used in encoding and decoding.



Figure 3. "barbara" original image

We use criterion, such as *PSNR*(Pick Signal to Noise Ratio) of formula 4 which represents original image or not and *bit rate* of formula 5 which can evaluate compression ratio according to image coding method proposed in this paper.

$$PSNR = 10 \log_{10} \frac{255^2}{\left\{ \sum \frac{(OPV - DPV)^2}{(nB)^2} \right\}} \quad (4)$$

*OPV*: Original Image Pixel Value

*DPV*: Decoded Image Pixel Value

*n*: block number, *B*: block size

$$bit \ rate = \frac{N_1 A_1 + N_2 A_2 + N_3 A_3}{(nB)^2} \quad (5)$$

*N<sub>1</sub>*: the Number of first step blocks

*N<sub>2</sub>*: the Number of second step blocks

*N<sub>3</sub>*: the Number of third step blocks

*I<sub>1</sub>*: Amount of first step block data

*I<sub>2</sub>*: Amount of second step block data

*I<sub>3</sub>*: Amount of third step block data



Table 2 shows that the encoding time of the method proposed in this paper can be reduced to about 99.82% of that of Jacquin's method, while the decoded image quality have above  $PSNR$  34[dB]. And the proposed method which performed limited domain searching method and block pattern code can reduce the 10% searching time compared with full searching wavelet domain, as the difference of the decoded image between the full searching method and the proposed method for each band is  $PSNR$  0.05[dB].

Table 2. Comparison of coding performance

	Jacquin's method	Full search & nocode	Proposed method Limited search & code
Enc. Time[s]	2327	40	4
	-	98.28%	99.82%
		-	10%
Ranges	16384	1423	1423
SNR[dB]	28.12	28.09	28.05
PSNR[dB]	34.58	34.35	34.30
Bit rate[bpp]	3.1	0.36	0.36
	-	88.38%	88.38%

The compression ratio of the method proposed in this paper was also enhanced compared to Jacquin's method using the original image without preprocessing. We can code image information of the band with high resolution for high compression ratio. This is the reason that the efficient image coding can be produced at the band with high energy resolution.

Figure 4 and Figure 5 shows each decoded image using Jacquin's method and the method proposed in this paper, which are almost similar.



Figure 4. Decoded image using Jacquin's



Figure 5. Our decoded image

## 5. CONCLUSION

In this paper, we propose the efficient image coding method by fast ITTBC algorithm using the 24 block pattern codes which are coded by the edge information in the image block on the subband segmentation, in order to improve the problem which the present ITTBCs have.

Compared with Jacquin's method, the proposed method in this paper can take the improvement of 99.82% searching time on condition that the decoded image quality should be maintained to  $PSNR$  34[dB], and the compression ratio can be improved to 88.38%. But it caused a little loss in quality of the decoded image because of the error by selecting the pattern of the blocks for the distinction of the similar blocks. We will enhance the image loss in quality if we improve the code for the block character.

We proved that the similar block searching time can be reduced by block pattern code and that the decoded image quality and the compression ratio can be improved simultaneously by adaptive multiresolution image partitioning on the subband segmentation.

## REFERENCE

1. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205-220, April 1992.
2. Benoit B. Mandelbrot, *Fractal Geometry of Nature*, Academic Press, 1988
3. Michael F. Barnsley, *Fractals Every where*, W H Freeman & Co, 1988
4. A. E. Jacquin, "Image Coding Based on a Fractal Theory of Iterated Contractive Image Transformation," *IEEE Transactions on Image Processing*, Vol. 1 No. 1, pp. 18 - 32, 1992.
5. Y. Fisher, *Fractal Image compression : Theroy and Application*, New York, 1994
6. M. Kawamata, ea. at. 1, "A Fast Coding Algorithm for Iterated Transformation Theory-Based Coding by Multi-

resolution Tree Searching," *The Journal of Electronic Information Communication Institute*, Vol, j78-A No. 2, pp. 253 - 260, 1995.

7. S. S. Koh, C. H. Kim, "A Fast Image Coding for Iterated Transformation Theory Based Coding Using Codebook and MRS(Multi-resolution Random Searching)," *ACCV'95*, Vol. 1 of 3, pp. I-434-I-438, 1995.
8. G. S. Cho, C. H. Kim, "Fractal Image Coding Method using Adaptive Pattern Code and MRS," *Basic Science and Engineering*, Vol. 1, No. 2, pp. 1029-1033, 1997.
9. T. A. Ramstad S. O. Aase, J. H. HusØy, *Subband Compression of Images : principles and examples*, ELSEVIER SCIENCE B. V., 1995.

# Current Research on the ARO-Positron Emission Tomography

Meei-Ling Jan\*, Hsing-Ching Liang, Shin-Wen Huang, Chuen-Shing Shyu  
Jiy-Shan Tang, Hong-Chih Liu, Cheng-Chih Pei, Ching-Kai Yeh

Physics Division, Institute of Nuclear Energy Research, Taiwan, ROC.

## ABSTRACT

We are presently constructing "AROPET", a rotating PET scanner for imaging small animals. The design of the system has flexible geometry, using four detectors. Each detector is made of a position-sensitive PMTs (Hamamatsu R3941) coupled with 18x16 small individual BGO scintillator crystals of dimension 2.6x2.6x25mm<sup>3</sup>. Animals can be imaged in two modes. One is similar to a gamma camera in which the detectors are stationary and a 2-D planar projection imaging is obtained. This mode is used for initial characterization of the bio-distribution of tracers. In the other mode the detectors are rotated through 90°, and the diameter can be adjusted between 22cm~40cm. This mode resembles a conventional 3-D PET scan using a partial detector ring. Thirty-one tomographic images can be obtained after rebinning and reconstruction. The field of view is ~1.3 mm (transaxial) by 45.6mm (axial). The spatial resolution of the planar projection mode, and the results of the planar image of a phantom and the dynamical images of the bio-distribution of F18-FDG in a mouse are discussed.

**Keywords:** Positron Emission Tomography, PET, planar projection imaging

## 1. INTRODUCTION

Animal models of human diseases are widely used in basic biomedical research to elucidate disease mechanism and to develop and test new treatments. Positron Emission Tomography (PET), an ideal, powerful tool in modern biology, allows the distribution of radiolabeled tracers of biological interest to be measured quantitatively and dynamically in living animals<sup>[1,2]</sup>. The ability to make repeated measurements on the same animal is unique to PET and confers an important advantage over traditional autoradiographic and tissue counting assays. In the basic research using experimental animals, research are interesting in the study of physiological process and the chemistry in living subjects. The animal PET experiments play an important role in basic research for bio-functions and in application studies such as the development of new drugs. However, PET systems developed for human used do not possess sufficient spatial resolution and sensitivity to accurately quantify changing organ radioactivity in small animals (mice, rats). Therefore, a number of groups have thought to overcome this limitation by developing high performance PET systems specifically for imaging small animals<sup>[3-8]</sup>. In this report, we describe an imaging system, ARO-PET (Animal ROTational PET), developed in our laboratory for the purpose of imaging mice and rats with two imaging modes, 2-D planar projection imaging and 3-D rotational tomography imaging. The scanner of the ARO-PET based on BGO crystals coupled to position-sensitive photomultiplier tube (PSPMT). The geometry of the scanner is flexible and allows adjusting the center-detector distance according to the sensitivity requirement and the size of the object being scanned.

## 2. SYSTEM DESCRIPTION

### A. Design Features

The design of the system has flexible geometry, using two pairs of detectors. The four detectors are mounted on a rotating plate which can be rotated by 90° during a scan by a stepper motor. The center-detector distance can be adjusted from 11cm~20cm. The configuration of the ARO-PET is shown in Figure 1.

The scanner has two emission imaging modes: 2-D planar and 3-D rotational (figure 2a, 2b). The 2-D planar mode produces planar projection images by using only one pair of detectors. From the detectors, signals in coincidence are readout and identified to two crystals by lookup tables. The planar imaging is performed by back-projecting lines of response joining these crystals. This mode will be valuable for measuring the distribution of radiotracer in an animal body as a function of time. The field-of-view (FOV) of the planar image is 51.5mm x 45.6mm.

\* Correspondences: Meei-Ling Jan Email: mljan@iner.gov.tw ; TEL: 886-3-4711400 ext.7403; FAX: 886-3-4711408

The rotational mode allows the four detectors to rotate around an object and obtains 3-D tomographic images. In this mode, the 3-D data mode will be rebinned to thirty-one sinograms, and be reconstructed by using 2-D Filtered-Backprojection algorithm<sup>[9]</sup>. This mode has the transaxial FOV 51.3 mm and the axial FOV 45.6mm.

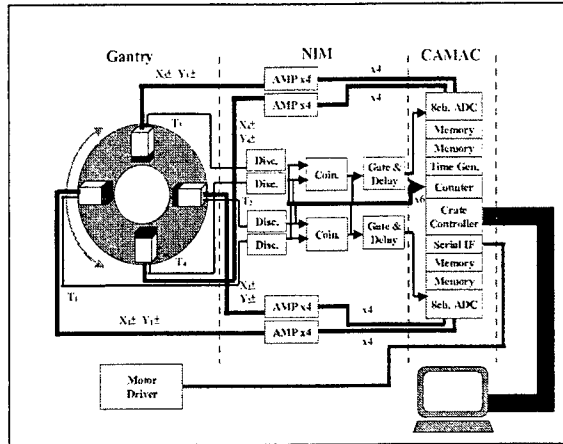


Figure 1: System schematic

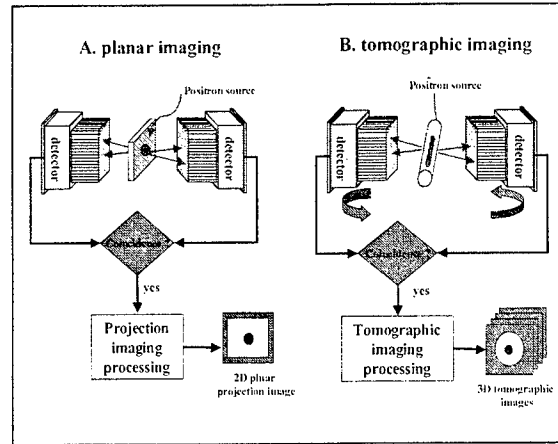


Figure 2: The ARO-PET has two imaging modes, planar imaging and rotational imaging. For rotational mode, all four detectors are used, although in this figure only two detectors are drawn.

## B. Detectors

One detector block consists of  $18 \times 16$  small individual BGO scintillator crystals of dimension  $2.6 \times 2.6 \times 25 \text{ mm}^3$ . All crystals are optically isolated by being wrapped with PTFE tape (figure 3). The crystal matrix is coupled to a PSPMT (Hamamatsu R3941). These detectors are mounted diagonally opposite to each other for detection pairs of the 511keV gamma rays. Since the two gamma rays produced from positron-electron annihilation are simultaneously emitted  $180^\circ$  apart, a logic AND module for timing coincidence is used to define the line of emission.

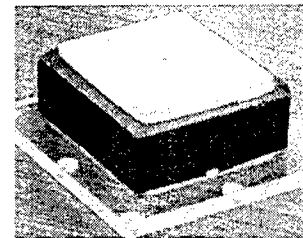


Figure 3:  $18 \times 16$  crystal matrix

## C. Data acquisition Electronics

The four signals, two for the x-direction and two for y-direction, from a PS-PMT are fed into a FERA ADC LeCroy 4300B for determining the x and y positions ( $x = (x1 - x2) / (x1 + x2)$ ,  $y = (y1 - y2) / (y1 + y2)$ ) and for pulse height (energy) analysis. The resulting four 11-bit data words are transferred by ECL-FERA bus to a FERA memory (LeCroy 4302 model) for temporally stored. After the FERA 4302 (32K words capacity) being full, all the data are transferred to a PC hard disk by CAMAC Dataway<sup>[10]</sup> via a PCI plug-in board KS2915 for advanced data processing. While a FERA memory is transferred data to PC, the other FERA memory replaces it to record the data from the ADC. Using two FERA memories to work in turn can minimize the dead time due to readout of the memory by CAMAC transfer. A timing signal from the last dynode of the PS-PMT is input a leading edge discriminator for suppressing low level noise, and then to a logic gate NIM (Nuclear Instrument Module) module LeCroy 365AL to determine whether the two detectors are truly in coincidence corresponding to a real positron annihilation. During a measurement, the LeCroy 365AL checks the coincidence condition with a coincident timing window 20ns and the ADC values represented the positions as well as the energy information are stored in list mode.

## D. Data Acquisition and Control Software

The data acquisition in 3D mode is performed by a PC, which also controls the stepper motors. After setting the scanned time and the number of rotating angle by an operator, the program starts to control the rotating plate and when the

plate has reached to its appropriate angle position, the program turns on the FERA ADCs and FERA memories<sup>[11]</sup>. During data acquisition, one of the 4302 memories is disabled (via CAMAC commands) and the ADC is enabled. Once the active memory asserts the LAM (Look-At-Me)<sup>[10]</sup>, the ADC is halted, the second memory is enabled, and then the ADC is cleared and enabled. While the data is being collected by the enabled memory, the data in the other memory is moved using block transfer to the PC. The process is repeated when the enabled memory asserts a LAM. Till the acquisition time has been achieved then the CAMAC modules are inhibit, and the data in the memory is transferred to the PC. After that the rotating plate rotates to the next angle, and the ADCs and the memories start again. The whole process is repeated until the detectors being rotated by 90°.

### 3. RESULTS

#### A. signals and effective counting rates Improvement

Position sensitive PMTs we used have a large Position dependent gain inhomogeneity, which leads to large position dependent variations in signal intensity. There are two problems of these detectors: (1) while the PSPMT's high voltage is not high enough (above the operation voltage 1000V already), parts of the crystals' responses are deficient. This is because some of the intensities of the signals can not exceed the discriminator's threshold, although the threshold level of discriminator has been adjusted as low as possible. This causes less the effective counting rate and sensitivity. (2) The other problem is that when the PSPMT's high voltage increases, the crystals' responses increase also, however some of them increase too many to be overflowed. Since the FERA 4300B ADC has its highest current limitation, the ADC value shows overflow while the intensity of signal from the detector exceeds the limitation. If one of the four signals from the same detector occurs overflow, this event will be not valuable for position and energy calculation. Therefore, the overflow of signal will cause the effective counting rate decreased. This can be seen in the table 1 and the figure 4. From the view of the energy spectrum, it shows in the figure 5 left, most of the overflowed signals should have the energy around the 511keV. So the loss of the overflowed signals will cause the coincident events with the primary energy to be lost. To avoid these problems, we adjusted the detectors' high voltage to 1250V, and used the home made attenuators with different factors for different detectors for reducing the signal intensity as well as the signal overflowed rate.

#### B. 2-D Planar Imaging

We used a phantom to obtain the 2-D planar imaging. All compartments of the phantom were filled with the F18-FDG solution. Figure 6 shows the planar projection image of this phantom. The scanned time was 5 minutes. Center-detector distance was 11cm. It is shown in figure 6, the upper part of the phantom image has higher activity, this is because the diameters of the upper part of the phantom are bigger. This is designed for the convenience of injecting in and drawing out the FDG solution from the phantom. Except the solution entrance and exit parts of phantom, the other parts are with the diameter of 2mm.

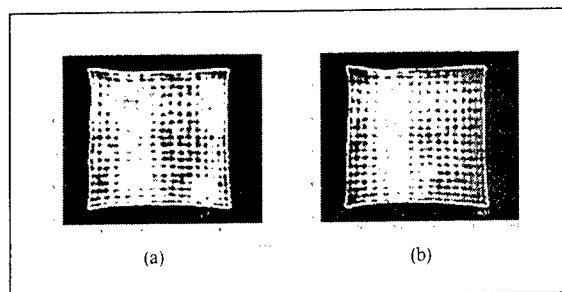


Figure 4: (a) before, and (b) after signal improvement. It shows in panel (b), the inhomogeneity of the crystal's response has been improved.

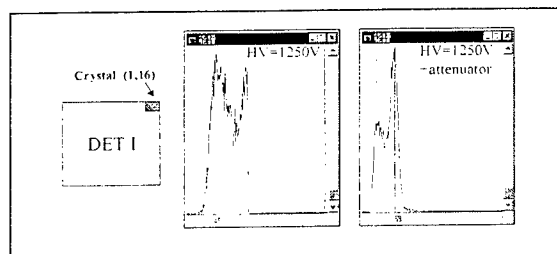


Figure 5: Energy spectrums of a crystal of the I detector. Left: most of the primary events are overflowed and can not be recorded. Right: More the primary events are collected by adding the attenuator to avoid the overflow of the signals.

The spatial resolution of the planar imaging of the ARO-PET was measured using a phantom with seven bars filled with F18-FDG. These bars are 4mm, 4mm, 3mm, 5mm, 5mm apart. The diameters of these bars are 2mm. Figure 7 shows the spatial resolution of the planar imaging of the ARO-PET system is 3mm or less.

First animal studies were performed also with the F18-FDG. The anaesthetized mouse was placed on the imaging bed, and it was injected with 0.5mCi F18-FDG. The center-Detector distance was 11cm. Scanning was started 10 minutes after injection for 60 minutes. The animal was viewed in face up with the nose at the top, the tail at the bottom. The lower half of the animal was in the field-of-view. Figure 8 shows the planar projection images of this experiment. The planar images of the 20mins, 40mins, 60mins after scanning being begun are shown. The brighter objects at the bottom of these images are the bladders. It shows that at the first 20mins only little of F18-FDG was accumulated at the bladder, but in the period of last 20mins there was the most F18-FDG being drained into the bladder, so the image of the bladder was the brightest. This experiment shows that the positron-emitting tracer can be dynamically followed with this device.

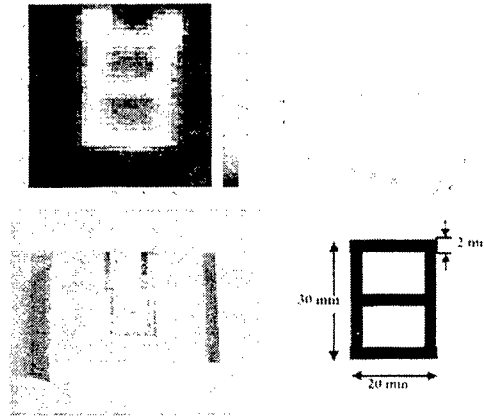


Figure 6: Planar image of a “田” phantom. The phantom was filled with F18-FDG.

#### 4. PERSPECTIVES AND DISCUSSION

The ARO-PET has been built over the last year in the Institute of Nuclear Energy Research. During fabrication it was realized the PSPMTs we got have defects in there. So the position response of the detectors were unsymmetry. After repairing the PSPMTs by the HAMAMATSU, we rebuilt the detectors and increased the crystal matrix to 18 x 16 to enlarge the FOV ( 51.3mm x 45.6mm). The signals from the detectors and effective counting rate are also improved. It is shown in this paper, the initial results obtained from the ARO-PET are encouraging. The planar imaging mode with spatial resolution less than 3mm can carry out a wide variety of studies with this device. For example, the dynamic planar projection image could be applied to time-activity curve measurements, and presents novel opportunities for modeling the transport of new radiopharmaceuticals and changes in organ function due to genetic or other manipulations. Besides, several experiments<sup>[11-13]</sup> showed the positron-based planar projection imaging technique can be applied to nonmedical applications, for example, to trace the uptake and transport of positron-emitting tracer in plants to observe the damage and recovery functions of plants *in vivo*.

We are currently working on measuring the rotating tomographic images. After further testing, we shall apply the tomographic mode of the ARO-PET system for phantoms and animal studies. Parameters such as the spatial and timing resolution, efficiency, and count-rate performance will be evaluated also.

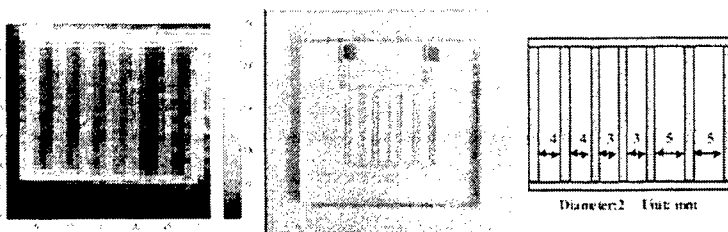


Figure 7: Planar image of a seven-bar phantom. The phantom was filled with F18-FDG. The seven bars are 4mm, 4mm, 3mm, 3mm, 5mm, 5mm apart.

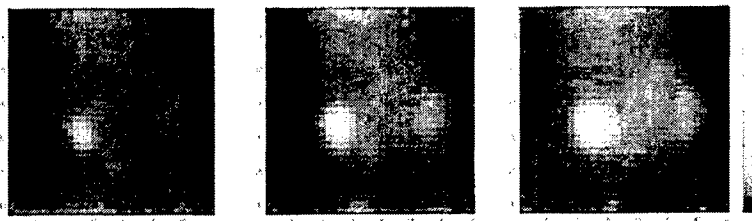


Figure 8: Planar images of a live mouse's lower half part. From left to right: sum image from 10~30mins, 30~50mins, 50~70mins after injection are shown. The bright objects are the bladder.

High Voltage	1150V	1175V	1200V	1250V
Effective Counting Rate (counts/sec)	2421	4632	6748	5388
Overflowed Rate(%)	1.7	1.9	4.1	22.8

Table 1: The effective counting rates and the overflowed rates of ADC values are varied with high voltage supply.

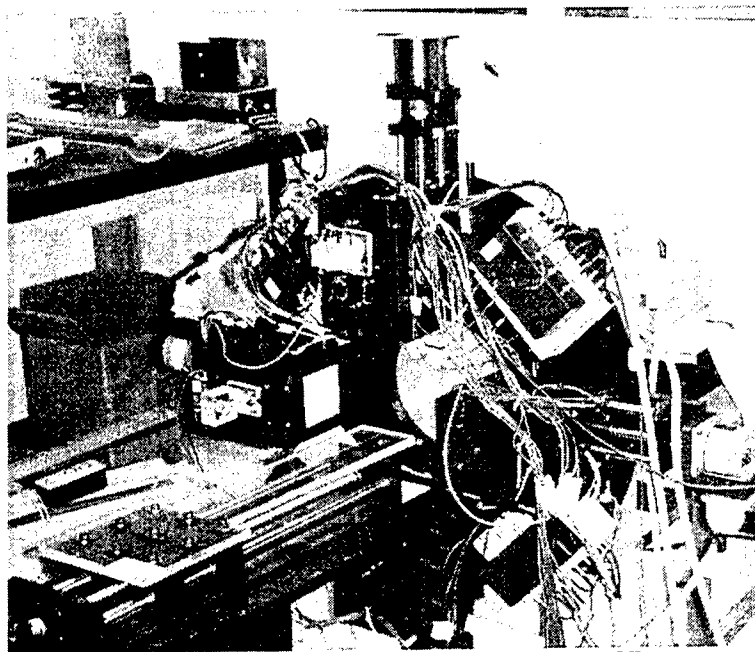


Figure 9: Photograph of the scanner of the ARO-PET system.

## 5. ACKNOWLEDGMENTS

The authors would like to thank the Isotope Applications Division of the Institute of Nuclear Energy Research, Taiwan, for supporting in producing the F18-FDG, and in preparing the anesthetized animals.

## 6. REFERENCE

1. M.P. Sandler, *et al.*, "Diagnostic Nuclear Medicine", Volume I, 3<sup>rd</sup> Edition, William & Wilkins, USA, 1996.
2. R.D. Hichwa, "Are animal scanners really necessary for PET?", *J. Nucl. Med.*, **35**, 1396-1397, 1994.
3. A.A. Lammertsma, "PET scanners for small animals", *J. Nucl. Med.*, **36**, 2391-2391, 1995.
4. S.R.Cherry, *et al.*, "MicroPET:a high resolution PET scanner for imaging small animals", *IEEE Trans. Nucl. Sci.*, **44**, 1161-1166, 1997.
5. M. Watanabe, *et al.*, "A high resolution PET for animal studies", *IEEE Trans. Imag.*, **11**, 577-580, 1992.
6. R. Lecomte, *et al.*, "Initial results from the Sherbrooke Avalanche Photodiode Positron Tomograph", *IEEE Trans. Nucl. Sci.*, **43**, 1952-1957, 1996.
7. S.Weber, *et al.*, "Evaluation of the TierPET system", *IEEE Trans. Nucl. Sci.*, **46**, 1177-1183, 1999.
8. S.Siegel, *et al.*, "Initial results from a PET/Planar small animal imaging system", *IEEE Trans. Nucl. Sci.*, **46**, 571-575, 1999.
9. R.A. Brooks, *et al.*, "Principles of computer assisted tomography (CAT) in radiographic and radioisotopic imaging", *Phys. Med. Biol.*, **21**, 689-732, 1976.



10. "An introduction to CAMAC", Lecroy 1994 Research Instrumentation Catalog.
11. T.K.Lewellen, *et al.*, "A data acquisition system for coincidence imaging using a conventional dual head gamma camera", *IEEE Nucl. Sci Symposium 7 Med. Img. Conference*, Nov. 9-15, Albuquerque, NM, USA, 1997
12. T. Kume, *et al.*, "Uptake and transport of positron-emitting tracer(<sup>18</sup>F) in plants", *Applied Radiation and Isotopes*, **48**, 1035-1043, 1997.
13. D.J. Parker, *et al.*, "Industrial positron-based imaging: principles and applications", *Nucl. Instr. and Meth. A*, **348**, 583-592, 1994.
14. G. M. Field, *et al.*, "Mechanics of powder mixing using positron emission tomography", *Nucl. Instr. and Meth. A*, **310**, 435-436, 1991.

# Influence of compression ratio of foam on printing quality of ink cartridge

Chin-Tai Chen<sup>\*a</sup>, Chien-Chang Lai<sup>\*b</sup>

<sup>a,b</sup>Printing Technology Division

Opto-Electronics & System Labs

Industrial Technology Research Institute

Bldg. 78,K120 OES/ITRI, Chutung 310, Taiwan, R.O.C.

## ABSTRACT

The paper had studied the influence of compression ratio of foam on printing quality of ink cartridge. Fundamental properties of foam were first introduced as a review. Thus, some basic models for compression of foam could have been built up to deal with the physical effect of compression on the back pressure of ink cartridge. It might actually make difference on many aspects for an ink-jet printer. By means of experiments, several individual cases had been done which composed of different combinations for the compressed foams and color inks. It turned out that the printing quality might be much influenced in the cases. The results implied that the compression ratio of foam should be correctly chosen in order to yield best quality on the print. Finally, detailed analysis explained to make a suggestion for the choice. It's expected to be very helpful in the future design of ink cartridge with the foam.

**Keywords:** Foam, Ink cartridge, Printing quality, Ink-jet printer, Porous material

## 1. INTRODUCTION

Porous material has been successfully applied in the manufacture of ink cartridge since about ten years ago<sup>1</sup>. Great progress for the application is still being made so far<sup>2,3,4</sup>. Compared with other types of ink cartridge<sup>5,6,7</sup>, the cartridge with porous material (so-called foam at most of time, in general) actually owned some specific advantages, such as high modulation of volume, good liability, low cost, and so on. Generally speaking, the manufacture of ink cartridge having porous material could be simply described in the state flow of Figure 1.0. It's noted that applying some compressive force onto the porous material could modulate the volume with ease. Hence, the compression ratio from one state to another could be achieved with no much difficulty of manufacturing flow.

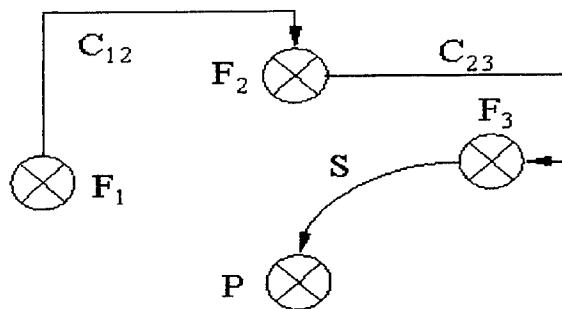


Fig. 1.0: State flow for manufacture of ink cartridge with porous material

<sup>\*a</sup> Correspondence: E-mail: chintai@itri.org.tw; Telephone: 886-3-5918358; Fax: 886-3-5917446

<sup>\*b</sup> Correspondence: E-mail: h880016@itri.org.tw; Telephone: 886-3-5916784; Fax: 886-3-5917446

The definition of symbols used in Figure 1.0 was also given on the Table 1.0. First of all, the raw foam  $F_1$  at original state could be compressed to a felted state of  $F_2$  by so-called felting method. The compression ratio would be defined as  $C_{12}$ . Secondly, the foam  $F_2$  would be loaded into a cartridge with compression ratio of  $C_{23}$ . Meanwhile, applicable ink would be filled into the foam  $F_3$  where the ink could be stored in the pores due to the capillary force. This capillary force was actually acting as a negative pressure to balance the gravity of ink in nature.

Foam's state :	$\begin{cases} F_1 \equiv \text{at original state} \\ F_2 \equiv \text{at felted state} \\ F_3 \equiv \text{at cartridge state} \end{cases}$
Compression ratio :	$\begin{cases} C_{12} \equiv \text{Compress } F_1 \text{ to } F_2 \\ C_{23} \equiv \text{Compress } F_2 \text{ to } F_3 \end{cases}$
Stored with ink :	$S \equiv \text{Surface tension of ink}$
Pressure state :	$P \equiv \text{Pressure in cartridge}$

Table. 1.0: Definition of symbols in the state flow

It's significantly noted here that the equilibrium of force could be expressed in the equation (1.a) where  $A$  was area of action,  $\rho$  was liquid density,  $g$  was acceleration of gravity,  $h$  was liquid height,  $L$  was length of action,  $\gamma$  was surface tension of liquid. The value of  $h$  could be exactly solved if a circular model of action with radii of  $r$  was assumed, as shown in equation (1.b).

$$A \times \rho g h \Big|_{\text{gravity}} = L \times \gamma \Big|_{\text{capillary-force}}, \quad (1.a)$$

$$h = \frac{2\pi \times \gamma}{\rho g r} \quad \text{if } A = \pi r^2, L = 2\pi r \quad (1.b)$$

Furthermore, two fundamental conditions should be at least satisfied in the ink cartridge. One was that the ink stored in cartridge could not leak throughout of ink outlet. It's defined as 'no leakage condition'. This condition meant that the capillary force of ink had to be greater than gravity of ink all the time in the foam. Mathematically, it's simply expressed in equation (2.a) where the symbol  $H$ , instead of  $h$ , represented maximum height of foam in cartridge. The other was that the ink could flow with ease to supply the need of nozzles in the print. No starvation of ink should occur in all print. It's defined as 'no starvation condition'. Mathematically, it could be described as equation (2.b) where the symbol  $f$  was introduced to represent the driving force as nozzle firing in the print.

$$A \times \rho g H \Big|_{\text{gravity}} < L \times \gamma \Big|_{\text{capillary-force}} \Leftrightarrow \text{no-leakage condition}, \quad (2.a)$$

$$A \times \rho g H \Big|_{\text{gravity}} + f \Big|_{\text{firing}} > L \times \gamma \Big|_{\text{capillary-force}} \Leftrightarrow \text{no-starvation condition}, \quad (2.b)$$

Additionally, the driving force might be created due to the vibration of PZT for the piezo-type of printhead; in the thermal bubble type of head, it could also be produced by the formation of bubble. Hence, the value of  $f$  was depending on what type of head would be applied to the ink cartridge. Experiments could figure it out. It would be further discussed later in next section.

### 3. COMPRESSION EFFECT

#### 3.1 Model of Compression

One 2-D model of compression was carefully considered as shown in Figure 3.0. Supposed that the porous foam was compressed by external force to change its shape from the initial state  $(A_1, L_1)$  to the final state of  $(A_2, L_2)$ . Noted that symbols  $A$  and  $L$  meant the total area and total perimeter of each pore (cell) in the foam. Meanwhile, for example, the four corners were transformed in the indicated paths ①, ②, ③, and ④, respectively. Thus, it's straightforward to define the compression ratio  $C$  as the ratio of  $A_1$  to  $A_2$ , i.e.  $C = A_1/A_2$ . It's important to see here that the total perimeter of each pore might not be changed in compression although the size of pore (cell) became smaller. In fact, we could make a assumption that cellular wall of each pore just got more curved to have no change of perimeter, i.e.  $L_1 = L_2$ . Hence, by applying the previous equation (1.a), it's predicted that the value of liquid height  $h$  would be multiplied by the  $C$  times.

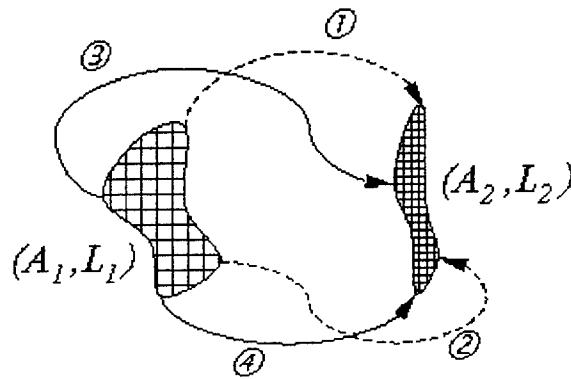


Fig 2.1: Physical model of compression

#### 3.2 Equivalent Principle

It's recalled that more than one time of compression might exist in the foam in the previous Figure 1.0. We'd like to solve such a series of compressions by presenting a simple 'equivalent principle'. The principle was illustrated in Figure 2.2. First, It's assumed that any compression made no difference at all. Therefore, the principle said that the final effect in a series of compressions would be equal to total multiplication of each one. Mathematically, the relationship could be exactly expressed in the equation (3).

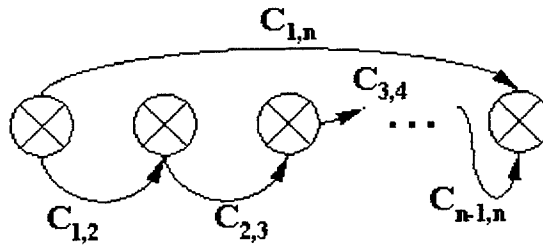


Fig 2.2: Equivalent principle of compression

$$C_{1,n} = C_{1,2} \times C_{2,3} \times C_{3,4} \dots \times C_{n-1,n} = \prod_{i=1}^{n-1} C_{i,i+1}, \quad (3)$$

## 4. EXPERIMENTS AND RESULTS

### 4.1 Regular Porous Material

Two different kinds of porous material (foam) were used to see if the 2-D model was working in the previous section 3.1. One kind was taken out from Epson cartridge and the other was from Canon cartridge. The pictures were shown in Figure 3.0. Left and right columns represented Epson and Canon, respectively; in addition, the four rows meant four different foams that were used to store four different colors of cyan, magenta, yellow, and black. In measurement of density, the Epson foams of  $0.06 \text{ g/cm}^3$  and Canon foams of  $0.13 \text{ g/cm}^3$  were obtained. It most possibly meant that the compression ratio  $C_{12}$  of Epson was greater than Canon's one. Thus, it was showing in the pictures that the Canon cells were curved more complicatedly than Epson ones. In fact, the foams in right column had greater capillary height than the left column did. Hence, the results could be a part of proof for the 2-D model.

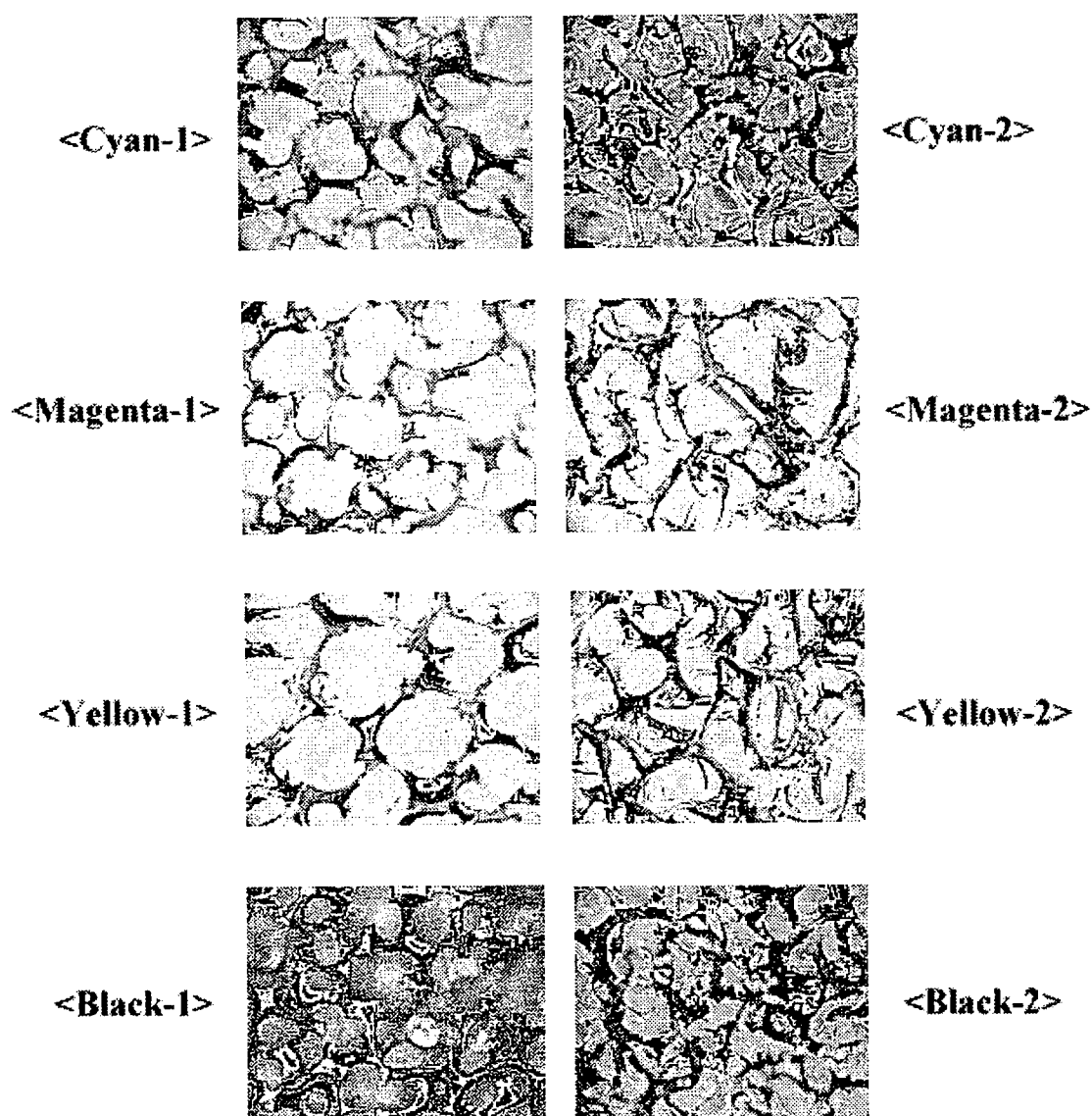


Fig 3.0: Pictures for two kinds of porous material

## 4.2 Linear Relation of Compression

Experiments were also done to see whether the relationship of compression effect was linear in 2D-model. First of all, four kinds of foam with different compression ratio  $C_{12}$  were prepared in the test, along with four different colors of inks. Noted that each piece of foam was  $10.1\text{mm} \times 16.0\text{mm} \times 45.7\text{mm}$  where the height  $H$  was  $45.7\text{mm}$  and the area  $A$  was  $10.1\text{mm} \times 16.0\text{mm} = 16\text{ cm}^2$ . However, in order to satisfy the need of capillary height, four pieces of foams were stacked upward together yielding total height of around  $18.3\text{ cm}$ . Next, the initial weight  $W_i$  (g) for each foam was measured as shown at the part <a> of Table 2.0. For convenience, all foams were storing with full level  $45.7\text{cm}$  of ink first and then put into a closed container for observation. It's obviously found that the ink level would automatically drop downward soon to its equilibrium state as described in equation (1.a). Total 47 hours were taken to make sure its stability of state. In the mean time, the final weight  $W_f$  (g) of each foam was measured then as shown at the part <b> of Table 2.0. Secondly, the void ratio V.R. of foam, defined as equation (4.a), was calculated as shown at the part <c> of Table 2.0. Finally, the equivalent height of capillary action could be figured out by applying the equation (4.b). The computational results of height (cm) were shown at the part <d> of Table 2.0.

$$V.R. = \frac{W_f - W_i|_{\text{no-ink}}}{V}, \quad (4.a)$$

$$h|_{\text{capillary-action}} = \frac{W_f - W_i}{A \times V.R.}, \quad (4.b)$$

<a>					<b>				
$C_{12}$	Yellow	Magenta	Cyan	Black	$C_{12}$	Yellow	Magenta	Cyan	Black
3	3.09	2.59	2.70	2.69	3	17.14	13.81	13.12	16.00
4	3.36	3.31	3.59	3.39	4	17.28	17.06	18.13	22.26
5	4.09	4.29	4.02	4.14	5	20.69	21.23	19.82	22.05
6	4.61	4.63	4.60	4.86	6	24.57	25.14	23.76	25.48

<c>				
$C_{12}$	$W_i$ (g)	$W_f$ (g)	$V_i$ ( $\text{cm}^3$ )	V.R. (%)
3	0.56	7.56	7.38	95
4	0.74	7.87	7.38	96
5	0.97	8.02	7.38	95
6	1.10	7.85	7.38	92

<d>				
$C_{12}$	Yellow	Magenta	Cyan	Black
3	9.15	7.30	6.79	8.67
4	8.97	8.86	9.37	12.16
5	10.81	11.03	10.30	11.67
6	13.43	13.80	12.89	13.87

Table 2.0: Experimental results of four foams for 2D-linear model, noting that  
 <a>initial weight of each foam before the test;  
 <b>final weight of each foam after the test;  
 <c>computation for void ratio of each foam;  
 <d>final capillary height of each foam after the 47-hour test.

Noted that the surface tensions of color ink (yellow, magenta, and cyan) were almost same of 35 dyne/cm; but black ink had a little higher value of 44 dyne/cm. To get rid of experimental error, it's preferred to choose the lowest height as a result in each row of part <d> of Table 2.0. If doing so, it's significantly found that the linear relationship held well on ratio of 3:4:5:6=6.79: 8.86: 11.03: 12.89. Hence, the results could have met well with the prediction of 2-D model as mentioned in previous section.

#### 4.3 Printing Quality

The experiments for influence of compression ratio of foam on printing quality were done as below. There were two different types of foams (Type A and B) chosen to be loaded into three cartridges (Cartridge A, B, and C). In the foam of Type A, two compression ratio of 2.5 and 2.75 were applied in the printing tests. Thus, total five pages of test pattern were printed. Finally, the results of printing quality was shown in the Table 3.1. Meanwhile, the foam Type B with two compression ratio of 2.5 and 3.0 were applied in the next printing tests, too. Once again, total five pages of test pattern were printed. Consequently, the results of printing quality was shown in the Table 3.2.

		Cartridge A	Cartridge B	Cartridge C
Foam Type A	Compression	<i>Printing Quality</i>		
	2.5	Excellent	Excellent	Excellent
	2.75	Excellent	Good	Bad

Table 3.1: Experimental results of print quality using foam type A with compression ratios of 2.5 and 2.75

		Cartridge A	Cartridge B	Cartridge C
Foam Type B	Compression	<i>Printing Quality</i>		
	2.5	Excellent	Excellent	Excellent
	3.0	Excellent	Excellent	Bad

Table 3. 2: Experimental results of print quality using foam type B with compression ratios of 2.5 and 3.0

It's noted that 'excellent' printing quality was defined as no any occurrence of banding lines in the five pages of print. Then, the 'good' printing quality was defined as the result with no more than two pages of banding in the tests. Otherwise, the printing result was defined as 'bad' printing quality in the tests. It seemed obviously that low compression would tend to yield better printing quality in the above test. It would be discussed for details later in the following section.

## 6. DISCUSSION AND CONCLUSION

### 6.1 Discussion for Results

The simple 2D-model had been examined with much care to determine two significant aspects of porous material in the study. By taking a lot of clear pictures of some regular foams, their microstructures showed that making more compression might cause more complicatedly curved cells. Since the change of total length for cells might be much less than the change of total area for the cells. Hence, the difference of capillary action could be predicted by the compression ratio of foam.

Secondly, the 'equivalent principle' was presented to further figure out the effects for a series of compressions that could be seen a lot in many ink cartridges. In order to see if it's accuracy, the linear relationship between compression ratio and capillary action should be proven. Some experiments were finished to find that it's approximately linear among the foams with different compression ratios. Noted that some judgements had been made to tell the relation because possible errors should be taken into account. In fact, it's pretty hard to measure the real height of capillary action. Therefore, some equivalent transformation was necessary in the experiments.

Finally, the tests for printing quality were finished. There were two types of foams and three different cartridges in the print tests. It's noted that no leakage must be satisfied in the first place, as shown in equation (2.a). Subsequently, the higher compression seemed to yield worse printing quality in the test cases. This could be explained by applying 'no starvation condition', as shown in equation (2.b). It probably meant that the high capillary force might cause the 'bad' printing quality. Firing force plus gravity for ink should be greater than capillary action so that excellent printing quality could be obtained.

### 6.2 Conclusion

The influence of compression ratio of foam on printing quality of ink cartridge was studied in the work. As a result, it was found that the compression truly played a significant role in the printing quality. Some key findings could be summarized as below-

- Simple 2D-model: it could successfully predict the effect of compression in advance. Actually, it's found that higher compression would yield more complexly curved cells of foam.
- Equivalent principle: it could be applied to calculate the total effect of a series of compressions. In fact, a linear relation was found among the effects of compressions. Experimental results had met well with the principle.
- No leakage and starvation conditions: the printing quality would be finally influenced by the two conditions. In fact, the conditions should be satisfied in the cartridge with compressed foam inside if excellent quality was desired.

Future work might be proceeding about the quantity of influence on printing quality in some degree. It's believed to be still much of interest in the near future<sup>8</sup>.

## ACKNOWLEDGEMENTS

This work had been supported by the program MOEA 894SY0000 for foam cartridge development project in Optics-Electronics System Labs of Industrial Technology Research Institute in Taiwan. The author really appreciated the support very much.

## REFERENCES

1. Baker et al, *Thermal Ink jet Pen Body Construction having Improved Ink Storage And Feed Capability*, US Patent 4771295, Hewlett-Packard Company, 1988.
2. Kotaki et al, *Ink Container Cartridge*, US Patent 5852457, Canon Kabushiki Kaisha, 1998.
3. Boyd et al, "Ink-Jet Pen With Near Net Size Porous Member," US Patent 5917527, Hewlett-Packard Company, 1999
4. Harshbarger et al, *Ink Cartridge With An Unfelted Foam And Method Of Printing*, US Patent 5892527, Lexmark International Inc., 1999.



5. Cowger et al, *Pressure-sensitive Accumulator For Ink-Jet Pens*, US Patent 5409134, Hewlett-Packard Corporation, 1995.
6. Baldwin et al, *Pressure Control Apparatus For An Ink Pen*, US Patent 5526030, Hewlett-Packard Company, 1996.
7. Khodapanah et al, *Ink Pressure Regulator for a Thermal Ink Jet Printer*, US Patent 5541632, Hewlett-Packard Company, 1996
8. Jali Heilman and Ulf Lindqvist, "The Effect of the Drop Size on the Print Quality in CIJ Printing," IS&T NIP15 Conference Proceeding, pp. 412-415, 1999.

# Influence of back pressure of ink cartridge on regular operation of ink supply system

Chin-Tai Chen\*

Printing Technology Division

Opto-Electronics & System Labs

Industrial Technology Research Institute

Bldg. 78, K120 OES/ITRI, Chutung 310, Taiwan, R.O.C.

## ABSTRACT

The influence of back pressure of ink cartridge on regular operation of ink supply system was dealt with in the paper. In recent years, large ink supply system has developed a lot to increase the print usage of cartridge; however, the amount of back pressure in the cartridge may change very much when the cartridge would be equipped with ink supply system. It might further result in different printing behavior of cartridge in the printer. It's much of interest in the study. Therefore, experiments were done to determine what influence might occur in one specific test system. It was found that ink droplet, nozzle firing, and print quality were significantly influenced over the time of ink supply. The experimental result was helpful in future design of ink supply system.

**Keywords:** Back pressure, Ink cartridge, Ink supply system, Ink-jet printer

## 1. INTRODUCTION

As the amount of print load on every ink-jet printer was getting larger, one large ink supply played a more important role than ever before. Meanwhile, several ink supply systems were already presented for the past years<sup>1, 2, 3</sup>. With these systems, the print usage of ink cartridge/print head could increase over a few times than ever before. However, some disadvantages related to the pressure stability of printhead might be induced<sup>4</sup>. It's recalled that the ink could be kept with no drilling in the printhead partly because a negative back pressure had to be successfully set up in the ink cartridge. Different back pressures might make difference of print thereof. Since the ink cartridge was connected to ink supply system, the three ones (cartridge, supply system, and print) could be related closely. The close relationship was illustrated in Figure 1.0 as follows.

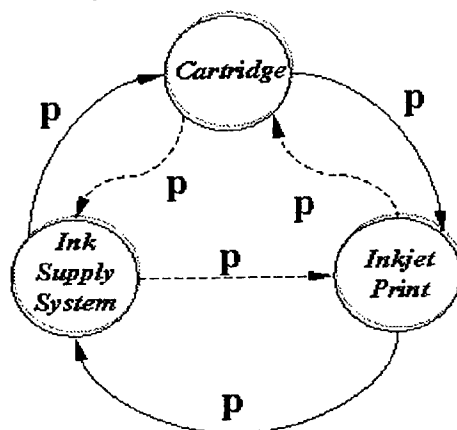


Fig. 1.0: Close relationship among ink cartridge, ink supply system, and inkjet print

\* Correspondence: E-mail: chintai@itri.org.tw; Telephone: 886-3-5918358; Fax: 886-3-5917446

We noticed that the pressure of ink, symbolized  $P$ , connected all of them together via any necessary tubing system. Any change of pressure in every one would create a signal of pressure wave to its neighbors in either clockwise direction (solid lines) or counterclockwise direction (dashed lines) of Figure 1.0. Therefore, it's expected to find the interactive effects in the study. In addition, we can carefully deal with the relationship by a simple mathematical model, as shown in Figure 2.0. Assumed here that any two among the ink cartridge, supply system, and inkjet print were symbolized of  $S_1$  and  $S_2$  (meaning the states). Next, there would be two possible statuses between  $S_1$  and  $S_2$ . One was 'standing status' that no ink was moving between the two states. In standing status, the pressures of two states must be equal, illustrated in equation (1). The other was assigned as 'flowing status' that ink was flowing from one state to another. No doubt, ink would always flow from high pressure state to low pressure state, expressed in equation (2). It's noticed that the flowing pressure loss should be taken into account mainly due to friction of motion.

$$P_{(S_1)} = P_{(S_2)} \Leftrightarrow \text{standing} , \quad (1)$$

$$\begin{cases} \text{if } P_{(S_1)} + P_{loss} > P_{(S_2)}, \text{ then } S_1 \rightarrow S_2 \\ \text{if } P_{(S_2)} + P_{loss} > P_{(S_1)}, \text{ then } S_2 \rightarrow S_1 \end{cases} \Leftrightarrow \text{flowing} , \quad (2)$$

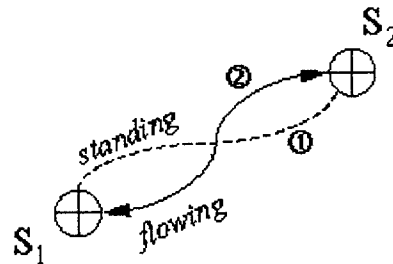


Fig. 2.0: Possible standing or flowing status between any two states

## 2. EXPERIMENTS

### 2.1 Test System Setup

One test system was set up to do the experiments, which composed of ink supply system, ink cartridge, etc., listed in the Table 1.0. Note that the cartridge was reconstructed for being able to obtain specific connection with the ink supply system.

① Ink	Pelikan Black Ink (water-based) with Surface tension = 48 dyne/cm, Viscosity = 2.35 cps = 2.35 x 10 <sup>-3</sup> N.S/m <sup>2</sup>
② Ink Supply System	◆ Rigid container with Maximum capacity = 375 cc Length x Width x Height = 62 x 46 x 134 mm
	◆ Tubing system (Silicone Tube) with Φ (outer) x Φ (inner) = 2 mm x 4 mm, Total length = 1700 mm
	◆ Guide (Flap Open and Close Type) with Total length = 900 mm ( 45 links, where one link = 20mm ) Width = 27 mm, Height = 12 mm
③ Ink Cartridge	HP51645A Length x Width x Height = 50 x 15 x 70 mm Maximum capacity = 53 ml, Weight with no ink = 53. g Weight with full ink = 106 g

Table 1.0: Key parts of test system in the experiments

## 2.2 Experimental Procedures

First of all, the ink cartridge was filled with full ink of volume  $V1$ . The reservoir of supply system contained ink of volume  $V2$  at the level of height  $H$ . Next, the cartridge was located at the vertical position where's lower  $h$  than that of the reservoir. The reservoir of system was simply open to the environmental air and yield  $P3$  of one atmosphere. It's critical here that the value of  $h$  should be less enough than the value of  $H$  to obtain a negative back pressure of  $P1$  and  $P2$  in the cartridge. Thus, the integrated test system could be running for print on the regular operation of inkjet printer. Secondly, having the system integrated, the experimental procedures were followed as shown Figure 3.0. It illustrated that the cartridge was driven to print with reservoir height of  $H$ ; of course, the  $H$  was decreasing over time. Then the reservoir would not be refilled to the full position  $H$  until the ink was low and almost running out. And repeated another print and stopped when the ink was low again.

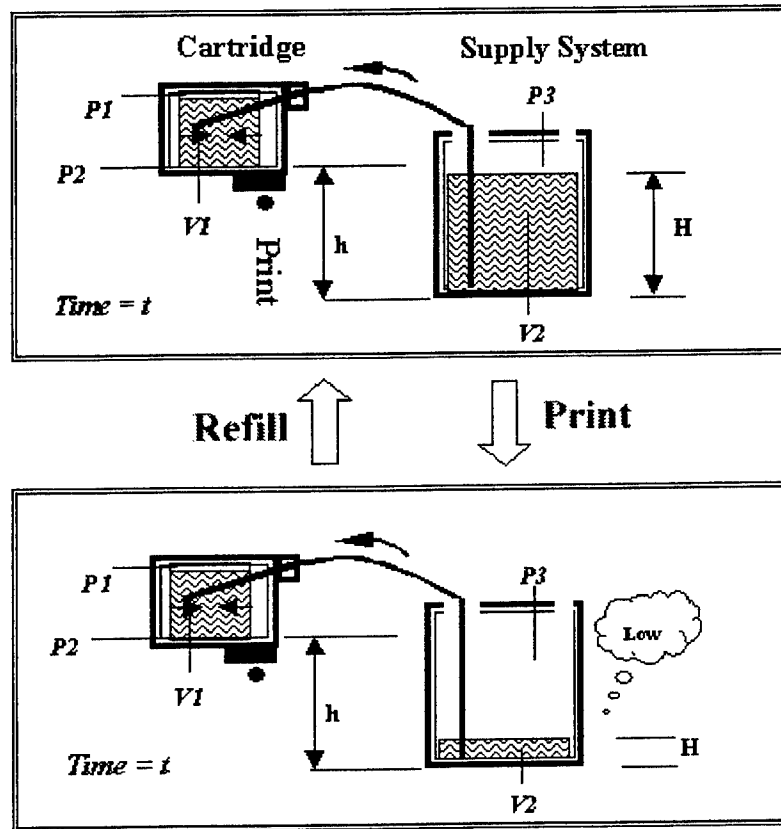


Fig. 3.0: Experimental procedures for the test system

Meanwhile, the initial conditions of the above experimental procedures were listed as below too. Consequently, all the results of print were measured and taken a record to find out the influence of back pressure on the operation of supply system. It's noted that all we concerned much would be the pressures  $P1$  and  $P2$  associated with the height of ink reservoir  $H$  in the system.

$$\left\{ \begin{array}{l} P3 = 1 \text{ atm} = \text{constant} \\ h = 12.3 \text{ cm} \\ H = 10.5 \text{ cm} \end{array} \right. \leftarrow \text{initial conditions at Time } t = 0$$

### 3. EXPERIMENTAL RESULTS

#### 3.1 Print History

The test had used A4 size of white paper as media in the print, which the printing pattern was simply a full-page black rectangle. As a result, total 680 pages (sheets) were printed on the two times of print and refill, noting that each time exactly finished 340 pages, respectively. The inking and pressure histories were recorded as below.

##### 3.1.1 Inking history

Inking history meant that the volume of ink left in ink reservoir, as be on printing, were measured once for every 20 pages. Therefore, the volume of ink versus page number could be simply taken a record as shown in Figure 4.1.

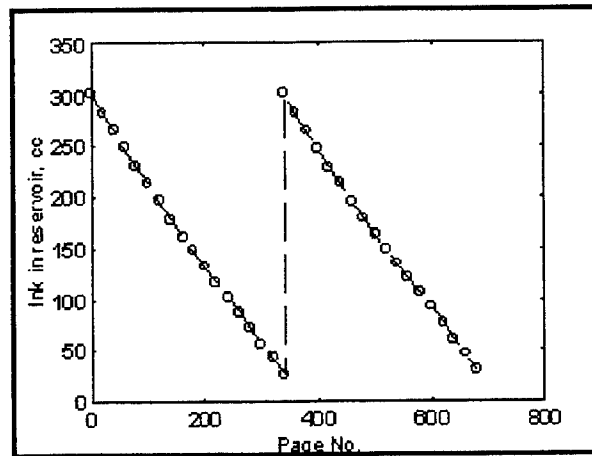


Fig. 4.1: Inking history of experiment

##### 3.1.2 Pressure history

Pressure history meant that the pressure status of ink cartridge, as be on printing, were measured once for every 20 pages too. Noted that pressure P1 of top position and pressure P2 of bottom position were necessary to be taken. As a result, the pressures versus page number could be recorded as illustrated in Figure 4.2.

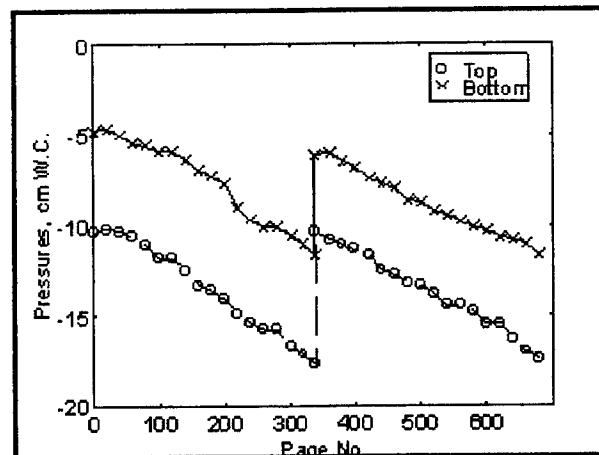


Fig. 4.2: Pressure history of experiment

### 3.2 Transformation to Change

The information of print history could be transformed into the change of corresponding properties. It's recalled here that what had be much concerned would be the situations of ink usage and pressures as the system working on every print. Hence, it's truly significant to make such a transformation as below.

#### 3.2.1 Change of ink usage

By applying the result of print history shown in Figure 4.1, the change of ink usage for every 20-page print could be calculated in the transformation. The result was illustrated in Figure 4.3. It was given with an average ink usage of 0.8 cc in the print history. In addition, with the printing resolution and coverage, an average drop size of ink was yielded of 36.72 pl. The two dashed lines A and B would be further discussed later in next sections.

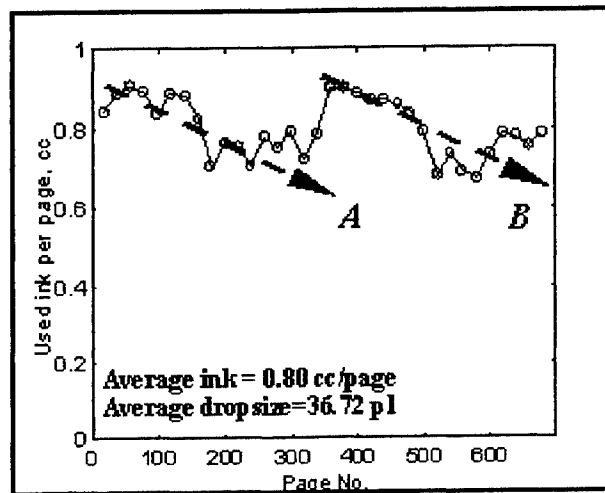


Fig. 4.3: Change of ink usage for the print

#### 3.2.2 Change of pressures

In the mean time, it's also concerned a lot that the relationship should be existing between the change of pressures in the reservoir and in the cartridge of head. By using the information in Figure 4.2, the relation was found as shown in Figure 4.4.

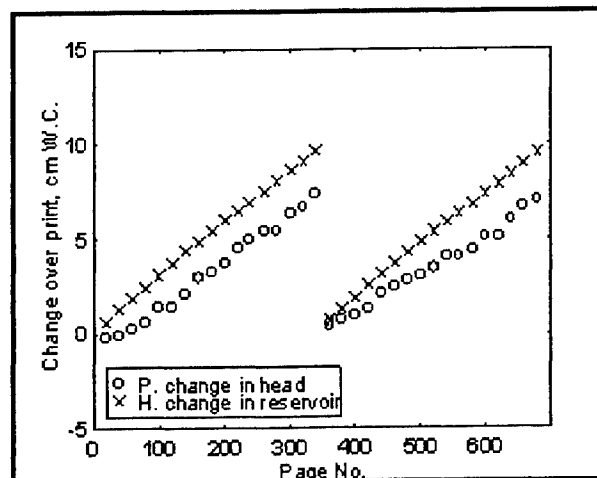


Fig. 4.4: Change of pressures in ink cartridge (head) and reservoir of supply system

### 3.3 Firing and Stability

It's reminded that the firing situation of nozzle and the stability of system might effect print quality. Simply printing one nozzle test pattern given in the test system of printer could check out the firing situation. Additionally, the stability of system could be defined as what change might be occurring during a long enough time.

#### 3.3.1 Firing situation of nozzle

Failure of nozzle (see symbol '\*' in Figure 4.5) might happen in the print. It had been checked by means of nozzle test print for every 100-page print. On the other hand, the equivalent amount of bad nozzles (see symbol 'x' in Figure 4.5) could be figured out in average way by transformation of Figure 4.3 where maximum usage represented 100% good nozzles and the rest was compared to it.

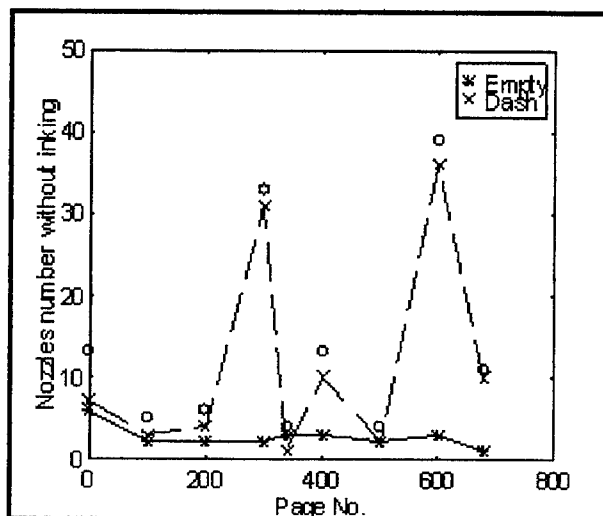


Fig. 4.5: Situation for failure and bad performance of nozzle in the print

#### 3.3.2 Stability of system

By above definition, the changes of weight in reservoir and pressure in head were recorded during a period of 12 hours, as shown in Figure 4.6.

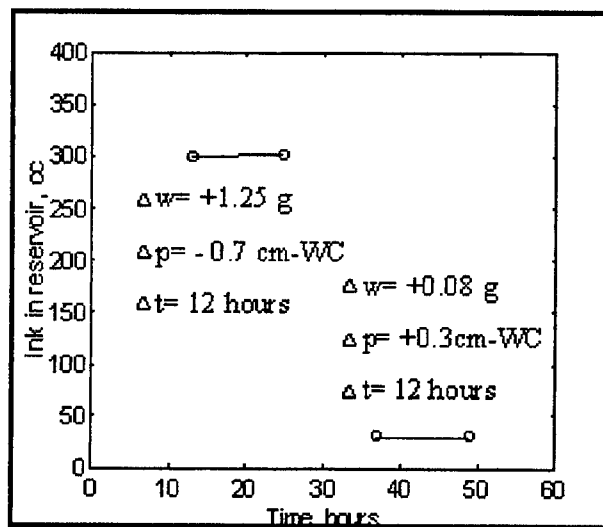


Fig. 4.6: Stability of system in the print

#### 4. DISCUSSION OF RESULTS

It's so obvious that the repeatability was pretty well in the two times of print and refill. This proved test system working fine. Thus, the influence of back pressure of ink cartridge on supply system had been seen clearly as follows. First of all, the back pressure of reservoir was simultaneously getting lower and lower as the pressure of ink cartridge was decreased (see Figure 4.2). The effect could be clearly explained in the previous equation (1). In physical point of view, the ink consumption of inkjet print sent a signal of pressure wave to cartridge and subsequently made a change of pressure in the cartridge. Following the pressure change of ink cartridge, the cartridge forwarded the signal to ink supply system and made a pressure change of supply system too. The Figure 1.0 could successfully explain this inside of experimental results. Additionally, the process would not stop until any two states of them had reached complete equilibrium, as described in Figure 2.0.

Secondly, the degree of effect could be defined in two aspects. One of them was to see how much change of pressure in cartridge might be caused by the change of pressure in reservoir. It's noted here that the cartridge HP51645A had a leaf spring acting as a pressure regulator. This regulator could eliminate some percentage of pressure change in reservoir. It's found in the experimental results that roughly 20-30% of pressure change might be eliminated. The other aspect was to check how much change of ink usage per page might be induced from the change of pressure in reservoir. Figure 4.3 had already obviously shown that the ink usage per page was decreasing from about 0.9 cc down to about 0.7 cc over time. The solid lines A and B in the Figure 4.3 clearly indicated the tendency in the direction of arrowhead. Noted that most of nozzles has no failure in the history (see Figure 4.5). Therefore, the tendency in decreasing might imply that the size of ink droplet had made difference over the time of print. The amount of change could be estimated in further calculation if the size changed. Of course, it could create some problems of print quality<sup>5</sup>.

Finally, the term of effect had been explored in the experiment. The experimental results had shown that any two states could keep quite stable during a period of 12 hours (see Figure 4.6). Of course, the state might take much less time than 12 hours to reach stable. If the period were shortened a lot, then one of the paths ① and ② shown in Figure 2.0 would make difference in the effect. It could be further explored in the future study.

#### 5. CONCLUSION

Influence of back pressure of ink cartridge on regular operation of ink supply system was studied in the paper. Physical concept for the influence was first presented and described in the aspects of pressure wave. Mathematically, some equations were given to predict the possible solutions of the effect. Thus, some experiments had been done to figure out the close relationship among the ink cartridge, ink supply system, and inkjet print. As a result, it's found that a signal of pressure wave should actually exist to connect each other. Consequently, the pressure change of ink cartridge had been simultaneously caused by the change of ink reservoir in the same way. In the mean time, the ink consumption per page of print had also been influenced with obvious tendency. More precisely, it's decreasing downward too when the pressure of cartridge was decreasing. Thus, some print quality might occur. The experimental results had met well with the prediction of physical models and mathematical equations.

In the future, further work could be explored on the effects of print quality, including ink droplet size, possible transformation in real time response of pressure, and so on. More interesting results in those aspects would be found and further help the future design of ink supply system.

#### ACKNOWLEDGEMENTS

This work had been supported by the program MOEA 883NB3110 for wide-format printer project in Optics-Electronics System Labs of Industrial Technology Research Institute in Taiwan<sup>6</sup>. The author really appreciated the support very much.

#### REFERENCES

1. Erickson et al, *Continuous Ink Refill System For Disposable Ink jet Cartridges Having A Predetermined Ink Capacity*, US Patent 5369429, LaserMaster Corporation, 1994.



2. Chuong C.Ta., *Rigid Tube Off-Axis Ink Supply*, US Patent 5691754, Hewlett-Packard Company, 1997.
3. Chin-Tai Chen, "Ink Tank Having Visible Ink Level and End Leg for Ink Supply System of Printer," IS&T NIP15 Conference Proceeding, pp. 59-61, 1999.
4. Seccombe et al, *Apparatus For Providing Ink To A Printhead*, US Patent 5650811, Hewlett-Packard Company, 1997
5. Jali Heilman and Ulf Lindqvist, "The Effect of the Drop Size on the Print Quality in CIJ Printing," IS&T NIP15 Conference Proceeding, pp. 412-415, 1999.
6. Chin-Tai Chen, *Design and Method for Large Ink Supply System(Chinese)*, OES-ITRI, Hsinchu, Taiwan R.O.C., 1999

# Method and apparatus for measuring the droplet frequency response of an ink jet printhead

Zhi-Ru Lian, Ming-Ling Lee, Yi-Hsuan Lai, Hung-Lien Hu, Chiehwen Wang

K200/OES/ITRI  
Bldg.78, 195-8, Sec.4, Chung Hsing Rd.,  
Chutung, Hsinchu 31040, Taiwan, R.O.C.

## ABSTRACT

To speed up the printing speed of an inkjet printer, the manufacturers normally focus on increasing the droplet frequency response. Hence, it has become a very important technique to measure the droplet frequency response of an inkjet printhead. A magneto-electric method is proposed to measure the droplet frequency response. The magneto-electric apparatus contains a metallic detecting plate and a magnetic ring with a gap of about 100 $\mu$ m filled with a nonmagnetic insulating material. The magnetic ring itself is made of a high-permeability alloy consisting of about 78% nickel and 22% iron. When an ink drop jetted from a nozzle makes a contact with the metallic detecting plate, which is perpendicular to the nozzle plate of a printhead, a current is conducted through the detecting plate immediately, and detected as a portion of expected signal. The expected signal is then processed by a signal processing circuit for counting the number of jetted drops, and determining the maximum droplet frequency response of the inkjet printhead as a function of the driving frequency of an applied voltage across the printhead.

Keywords: Frequency response, magneto-electric method

## INTRODUCTION

For most commercial inkjet printers, printing graphics and documents is normally carried out by the printhead. In principle, a thermal bubble printhead of an inkjet printer heats up the ink and vaporizes the ink to form ink bubbles by converting electric energy into heat. The printhead then jets the ink drops, which are developed from the ink bubbles, onto a medium surface through spouts. In order to speed up the printing efficiency of an inkjet printer, the manufacturers normally focus on increasing the droplet frequency response. That is, the droplet frequency response indicates the printing speed of an inkjet printer. Hence, how to measure the droplet frequency response of an inkjet printhead has become a very important technique in inkjet printer manufacture.

## MEASUREMENT METHOD AND EXPERIMENTAL APPARATUS

The droplet frequency response is obtained by comparing the detected actual jetting frequency of an inkjet printhead with the driving frequency actually applied to the inkjet head. The maximum droplet frequency response of the inkjet printhead can be measured by checking the matching between different driving frequencies and the actual responding jetting frequencies. Since the ink bubbles are generated from the printhead in a frequency varied from several kilo-Hertz (kHz) to several tens kHz, it is impossible to detect the actual droplet frequency response through a regular image snapping system. Even though utilizing a high-speed camera is capable of catching the actual droplet frequency response of an inkjet printhead, and then, to determine the droplet frequency response of the inkjet printhead. However, it is not cost effective. Hence, some apparatuses and methods have been developed for the purpose of measuring droplet frequency response of an inkjet printhead, such as those disclosed by US patent number 4,484,199<sup>(1)</sup> and US patent number 4,590,482<sup>(2)</sup>.

The schematic cross-sectional diagram of a conventional measuring apparatus for determining the droplet frequency response is illustrated in Fig.1. As seen from Fig.1, a planar detecting electrode is placed parallel to a metallic nozzle plate,

and a voltage is applied across the detecting electrode and the nozzle plate. The detecting electrode and the nozzle plate are not electrically connected, though the distance between them is quite short. The distance is less than 100  $\mu\text{m}$ . Once an ink drop is jetted by the nozzle plate through nozzle, the ink drop forms an electric conduction between the detecting electrode and the nozzle plate before the ink drop totally leaves the nozzle plate. A series of electric conduction formed by continuously jetted ink drops out of the nozzle plate can be detected by an attached electronic circuit (not shown in figure) for obtaining the forming frequency of ink drops. However, ink drops are easily stuck within the narrow space between the detecting electrode and the nozzle plate, and that leads to an error reading on the forming frequency of ink drops while a detecting process is performed.

The schematic cross-sectional diagram of another conventional measuring apparatus for determining the droplet frequency response is illustrated in Fig.2. Referring to Fig.2, a pair of electrodes is placed between the nozzle plate and the detecting electrode, wherein a high voltage is applied across the electrodes to provide a high-voltage electric field. While an ink drop jetted by the nozzle plate passes through the electrodes, the ink drop is charged. An electric signal can then be detected at the detecting electrode after the charged ink drop hits the detecting electrode. By counting the number of the electric signals within a period of time, the forming frequency of the ink drops is obtained. An ink drop, which is about 100 pico liters (pl) in volume, is possibly broken into several sub-drops while the ink drop passes through the high-voltage electric field, says exceeding 1000 volts. Therefore, the detected frequency counting at the detecting electrode is interfered by the noise signals given by the sub-drops.

A magneto-electric method is proposed to ensure a more precise measurement on the droplet frequency response, which does not encounter the problems of noise signal and error reading as mentioned previously. The magneto-electric apparatus for measuring the frequency response of ink drops contains a metallic detecting plate and a magnetic ring both placed under the nozzle plate, as shown in Fig.3A and Fig.3B. The detecting plate is perpendicular to the nozzle plate. In order to prevent an erroneous reading caused by stuck ink drops gathering on the detecting plate, the lower section of the detecting plate is designed to be capable of draining ink drops efficiently. Since the ink drops are formed at a pretty high forming frequency, from several kHz to several tens kHz, an erroneous reading is possibly obtained if the measured ink drops can not be efficiently drained. According to the foregoing consideration, the lower section of the detecting plate, for example, is made to be a metallic net-like structure, or a plate with a sharp corner pointing downward as shown in Fig.4. With a net-like structure or a sharp-corner shape, ink drops dropped on the detecting plate tend toward getting together as a larger drop, which is easily drained from the detecting plate.

As seen from Fig.3B, the magnetic ring has an opening toward the detecting plate, wherein the magnetic ring is attached to the detecting plate with the side arms aside the opening. The plane circled by the magnetic ring is perpendicular to the detecting plate, and parallel to the ground. The magnetic ring is, for example, an about 0.3-mm-thick lamination consisting of high-permeability material films or high-permeability alloy films. The selected high-permeability alloy can be an alloy of about 78% nickel and about 22% iron or other alloys with the similar properties. The selected high-permeability material can be ferrite, or other materials with the similar properties. The air gap of the magnetic ring is about 100 to 150  $\mu\text{m}$ .

As seen from Fig.4 and Fig.3A, an insulating layer is placed between the nozzle plate and the detecting plate to prevent unnecessary electric conduction between the nozzle plate and the detecting plate. The insulating layer is about tens of microns to 100  $\mu\text{m}$  in thickness. While a detecting task is performed, the measuring apparatus consisting of the insulating layer, the magnetic ring and the detecting plate is moving along the nozzle plate. The insulating layer is also used here to ensure the minimum distance between the detecting plate and the nozzle plate is fixed to a pre-determined distance, about the thickness of the insulating layer. The distance between the nozzle plate and the detecting plate can be reasonably adjusted and has to be short enough, so that an ink drop jetted from the nozzle can still make an electric conduction between those two plates before it drop off from the nozzle plate. All detected electric signals are output through a signal wire, which is electrically connected to the detecting plate, to a signal processor (not shown in figure). The measuring apparatus also contains a holding apparatus (see Fig.3A), and a supporting arm (see Fig.4). The holding apparatus is used to hold the magnetic ring, and the supporting arm is used to support and move the entire measuring apparatus.

The method for measuring the droplet frequency response by utilizing the foregoing measuring apparatus is based on the magneto-electric principle. As shown in Fig.3A, once a detecting task is started, a voltage is applied to the nozzle plate through a probe (not shown in figure). The voltage is about 30 volts and is capable of providing a current that is no higher than 100mA while a close loop is formed. When an ink drop is jetted from the nozzle, before the ink drop totally drops off from the nozzle, it forms an electric conduction between the nozzle plate and the detecting plate. As a result, a current  $I$  then flows through the detecting plate.

According to the Lenz's law, an induced magnetic field, which relates to the variation of current, is then generated by the formation of current  $I$  flowing through the detecting plate. Since the direction of current  $I$  is parallel to the detecting plate, the magnetic lines of force of the induced magnetic field generated by the show-up of the current  $I$  are perpendicular to the detecting plate. Therefore, the magnetic ring has to be placed in the position that the area circled thereby is perpendicular to the detecting plate in order to sense the induced magnetic field.

As soon as the ink drop totally drops off from the nozzle, an induced current  $I'$  flowing in the same direction as the current  $I$  is generated by the magnetic ring accordingly to the Lenz's law. Through the signal wire, the variation of voltage and current over the detecting plate within a time frame is fed to a signal processing routine (not shown in figure) to be further processed.

## EXPERIMENTAL RESULTS AND DISCUSSION

The waveform of a detected electric signal is illustrated in Fig.5. The x-axis represents time and the y-axis represents the voltage of the detected electric signal at a corresponding time. The detected electric signal includes two segments, a fore-signal happening within the time frame  $T1$  and a post-signal happening within the time frame  $T2$ , wherein the fore-signal is corresponding to the closed-loop current  $I$ , and the post-signal is then corresponding to the induced current  $I'$ . The time frame starts at when the ink drop jetted by the nozzle begins to make a contact with the detecting plate, wherein a portion of the ink drop, contacting interface, is connected to the detecting plate while a contact is made. The area of the contacting interface is increased within the time frame, and reaches its maximum at the end of the time frame, that is, the ink drop is dropped off completely from the nozzle. The post-signal detected within the time frame is the induced current  $I'$  generated by the magnetic ring due to the variation of current on the detecting plate. The induced current  $I'$  flows in the same direction as the closed loop current  $I$  does, and is gradually decreased as time goes. Without the presence of the magnetic ring, the only signal detected is the narrow and sharp pulse as shown in the time frame of Fig.5 that is difficult to detect. Therefore, the measuring apparatus increases the sensitivity of the measurement by adding a magnetic ring. While the printhead is operating by applying a driving signal, every ink drop jetted from the nozzle gives an electric signal detected by the magneto-electric measuring apparatus as shown in Fig.5.

A experimental result showing the electric signals detected by the measuring apparatus within a period of time is recorded as shown in Fig.6, wherein the x-axis represents time and the y-axis represents the voltage. The waveform signal in Fig.6 can be further processed to obtain a number indicating the forming frequency of ink drops at the nozzle plate. By checking the degrees of match between the forming frequencies of ink drops and the corresponding driving frequencies, the maximum droplet frequency response of the printhead of an inkjet printer can be obtained. The electric signals obtained on the detecting plate are sent to a signal-processing routine, and processed in a manner as shown in Fig.7.

After the electric signals are fed into the signal-processing routine through signal wire, a signal processor then picks up the valid signals first. The valid signals are next further adjusted and cleared by using a filter and a corrector to eliminate the noise signal. The results are digitized into digital signals. By using a display, such as a monitor, the digital signals are displayed on the monitor in the format of a waveform. Then, by checking the matching degrees of pairs of waveforms, each pair of waveforms consists of the forming frequency of ink drops and the corresponding driving frequency. The maximum droplet frequency response of the inkjet printhead is then obtained.

The insulating layer of the measuring apparatus prevent undesired electric conduction between the detecting plate and the nozzle plate, so the erroneous reading caused by improper electric conduction is avoided. The detecting plate perpendicular to the nozzle plate is capable of draining the dropped ink drops efficiently, so that no ink drop is stuck between the detecting plate and the nozzle plate that affect the detected results.

The magnetic ring of the measuring apparatus further enhances the detected signals, so the detected results are more easily to be processed for obtaining more precise results.

## CONCLUSION

Based upon Lenz's law, a magneto-electric apparatus is designed to measure the frequency response of ink drops. The measurement sensitivity of the apparatus is drastically increases by adding a magnetic ring as compared to some conventional apparatuses. The magnetic ring further enhances the detected signals by generating an induced current, so the detected results are more easily to be processed for obtaining more precise results.

## REFERENCES

1. Masato Watanabe, "Method and apparatus for detecting failure of an ink jet printing device," US patent No.473654 (1984).
2. Robert R. Hay and Paul R. Spencer, "Nozzle test apparatus and method for thermal ink jet system," US patent No.4590482 (1986).

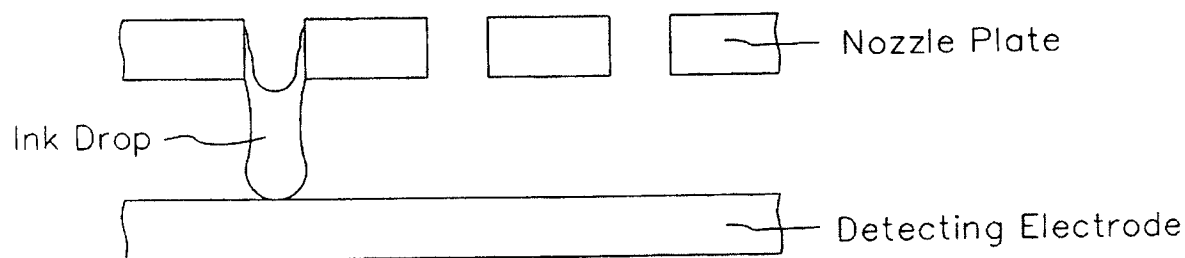


Fig.1 A schematic side-viewed diagram showing a conventional measuring apparatus for detecting the forming frequency of ink drops.

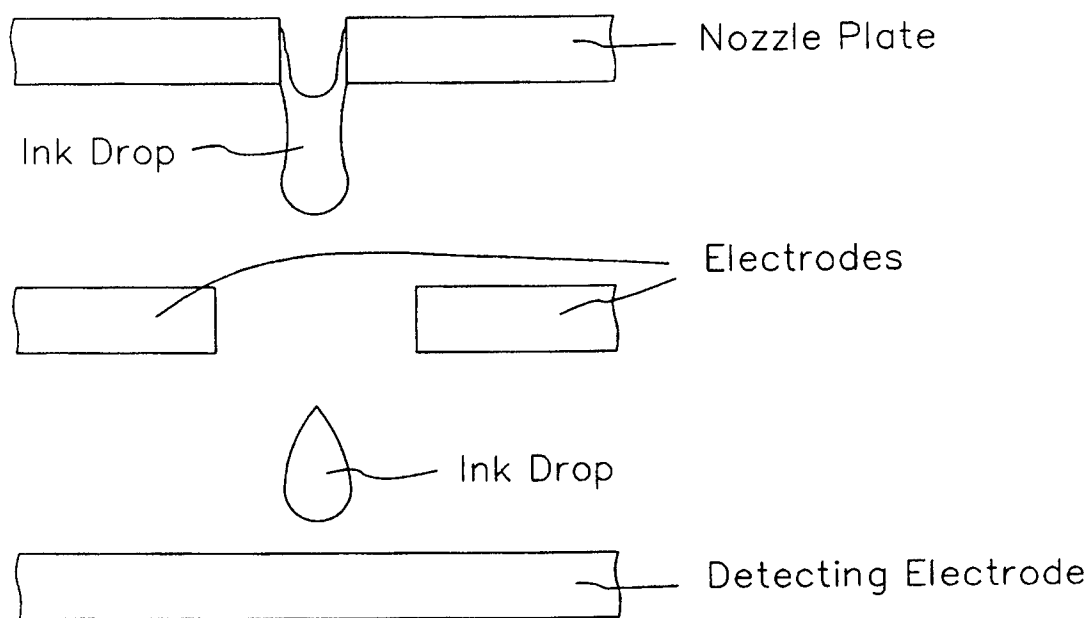
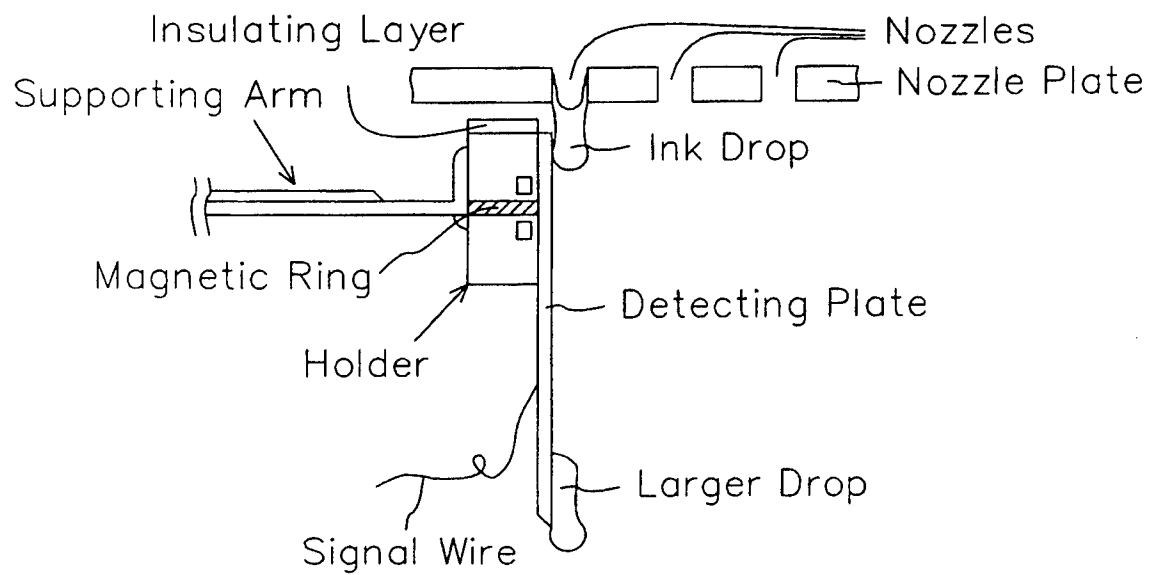
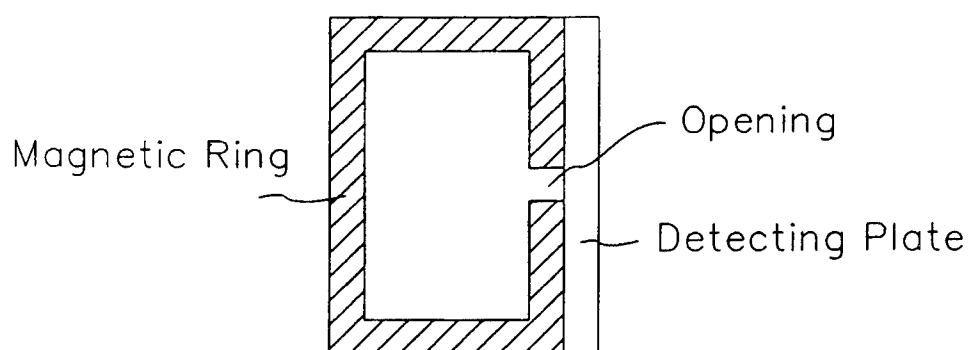


Fig.2 A schematic side-viewed diagram showing another conventional measuring apparatus for detecting the forming frequency of ink drops.



(3A)



(3B)

Fig.3 (A) A measuring apparatus for detecting the forming frequency of ink drops. (B) Magnetic ring

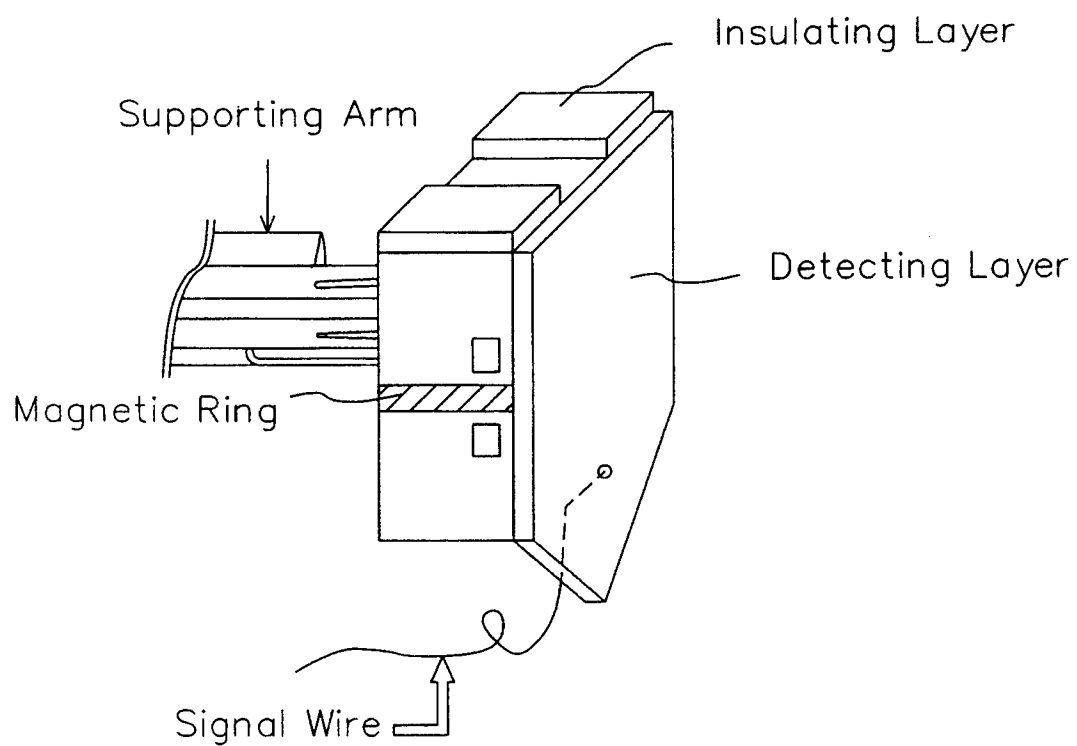


Fig.4 A schematic top-viewed diagram showing a measuring apparatus for detecting the forming frequency of ink drops.

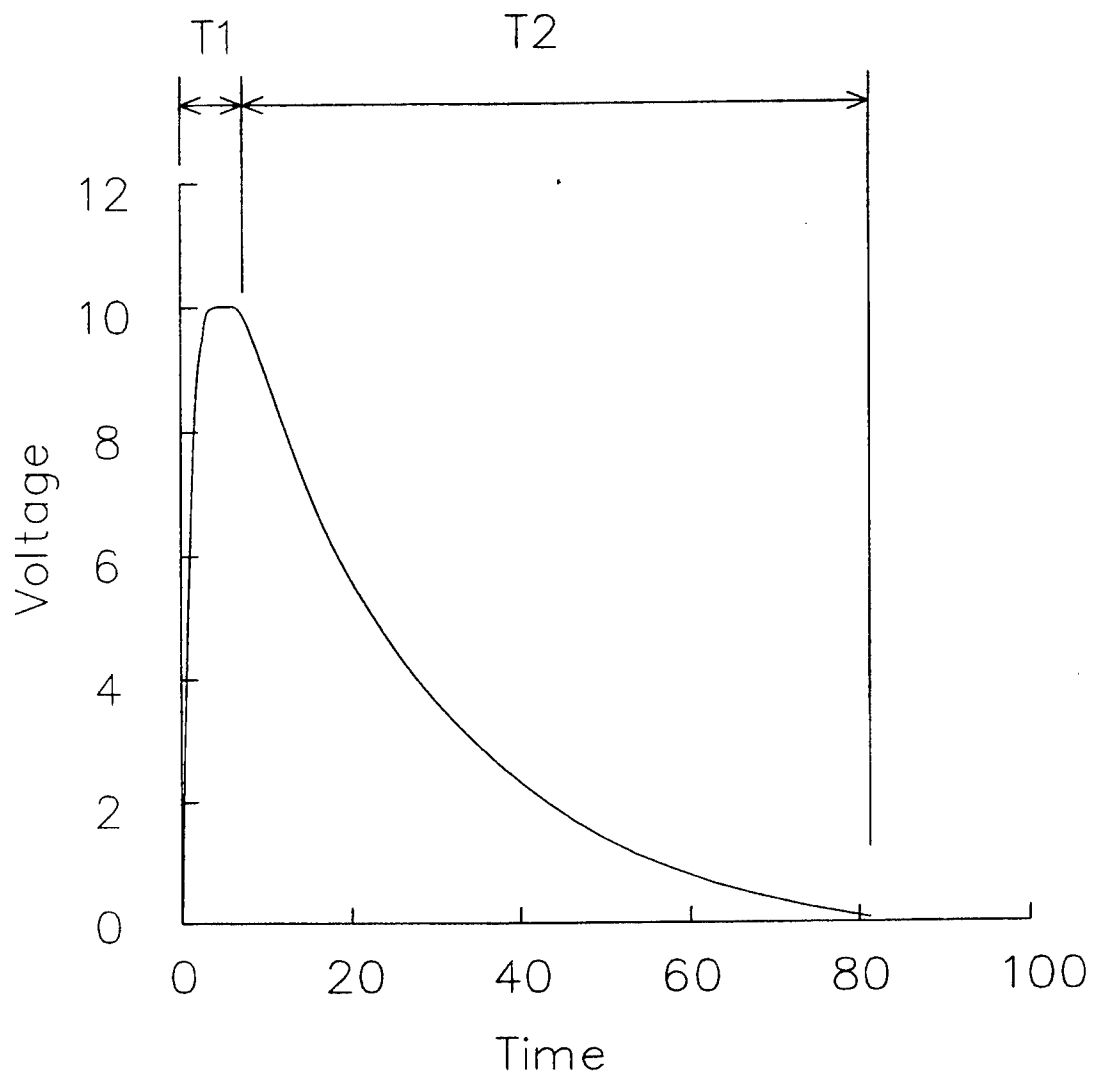


Fig.5 a waveform plot showing a signal detected by the measuring apparatus for detecting the forming frequency of ink drops. Time frame T1 Corresponding to the closed-loop current  $I$ . Time frame T2 corresponding to the induced current  $I'$ .



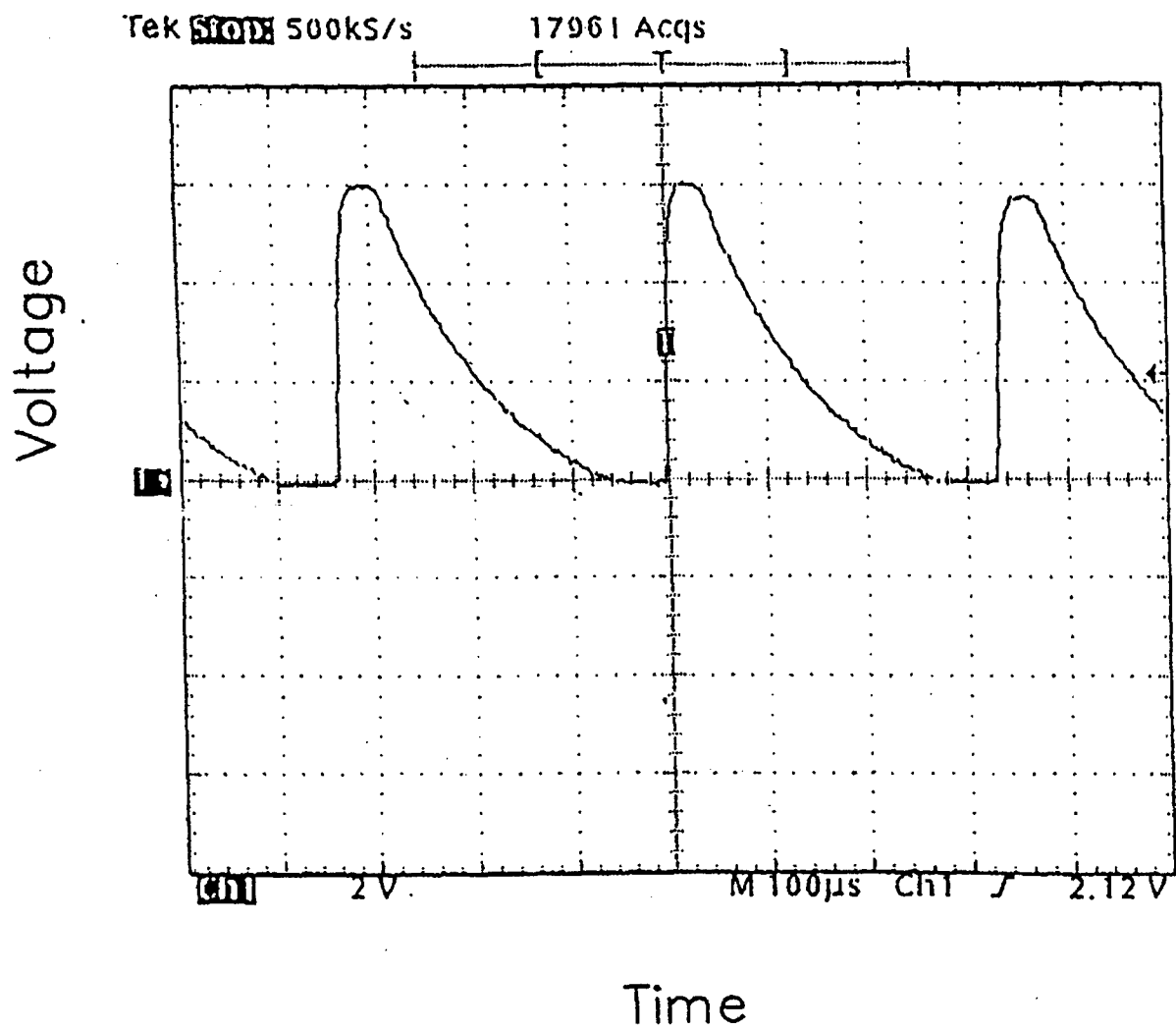


Fig.6 The actual signal measured by the magneto-electric apparatus for detecting the forming frequency of ink drops.

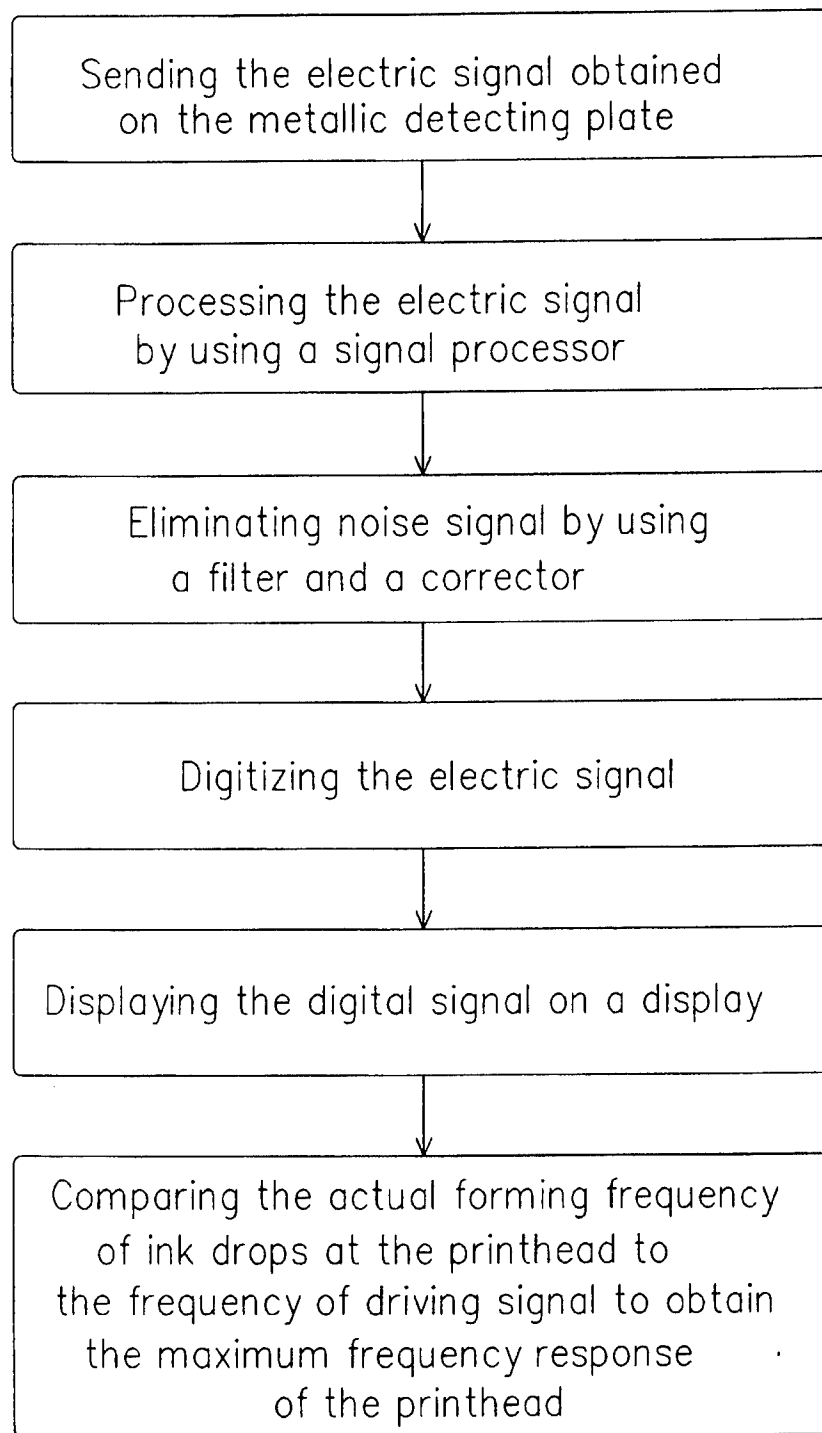


Fig.7 The flowchart of signal-processing routine used to process the signals detected by the measuring apparatus.

# Film Stress and Adhesion Characteristics of Passivation layers for Thermal ink-jet printhead

Yih-Shing Lee <sup>a</sup>, Yi-Yung Wu<sup>a</sup>, Chen-Yue Cheng<sup>a</sup> and D. S. Wu<sup>b</sup>

<sup>a</sup>Opto-Electronics & Systems Laboratories, ITRI

<sup>b</sup>Electrical Engineering department, Da-Yeh University

## ABSTRACT

Amorphous, hydrogenated silicon carbide (a-SiC:H) deposited by plasma-enhanced-chemical-vapor-deposition (PECVD) has been used as the most important film of passivation layers in a thermal ink-jet printhead. When the printhead was thermal-cycled from room temperature to about 400°C, the a-SiC : H film is sustained by a variety of thermal and mechanical stresses that are detrimental to its integrity. Thermal stress changes of a-SiC:H films were varied with different CH<sub>4</sub>/SiH<sub>4</sub> gas ratios. Microstructure investigation was mainly achieved by FTIR technique. Less Variation of the Si-H absorption bond causes less thermal stress change.

Thin-film adhesion is an important problem in thermal ink-jet printhead between the Ta thin film and a-SiC:H films. A qualitative measure of film adhesion can be made with the scratch tester. The adhesive critical load and Ta coating failure modes on a-SiC:H were acquired to examine the film adhesion of these two investigated films. The adhesion depends on the nature of the interfacial region, which depends on the interactions between the depositing Ta thin film and the surface a-SiC : H films. An increased effective contact area in the interfacial region promotes a good adhesion.

**Keyword:** Thermal ink-jet printhead, a-SiC:H, thermal stress changes, film adhesion

## 1. INTRODUCTION

The printhead resistor structure for thermally exciting the ink ejection is fabricated on a silicon (Si) substrate using standard IC processing steps. <sup>1</sup> Silicon oxide is deposited on the Si substrate as a barrier layer to prevent leaching of impurities. The resistor and conductive layers are tantalum (Ta)-aluminum (Al) and Al, respectively. The resistor-conductor films are lithographically patterned to form individual heater and conductive stripe line. Then heater and conductor are covered with ink-resistant passivation films. A dry-film coating further protects the passivation and the underlying thin films from degradation by the ink. To improve contact reliability, the Al pads are coated with Ta and gold (Au) films. A thin film resistor film is rapidly heated through a joule-heating process by direct-current (d.c.) electrical pulses which pass through the resistor and each time last a few microseconds. <sup>2</sup> The temperature the top surface of the device is raised, instantaneously, to over 300°C. The ink-vapor bubble formed adjacent to the resistor propels an ink droplet out of nozzle to form a dot on the paper. After the electrical pulse is turned off, the vapor bubble collapses, subjecting the thin-film passivation to severe hydraulic forces. During the operation and life of the printhead, the passivation experiences severe electrical, thermal, mechanical, and chemical stresses. Developing a passivation film for these exacting requirements presented some interesting challenges. In the present printhead structure, in its operation mode, is thermally cycled between room temperature and 300-plus °C, at repetition rates of several kHz. In this respect, thermal stress in each layer due to thermal expansion mismatch is calculated. Chang et al. <sup>2</sup> have showed that stresses in the carbide (SiC) and resistor layers due to thermal expansion mismatch may approach 10<sup>10</sup> dyne/cm<sup>2</sup> under normal condition. Therefore, thermal stress changes of SiC film play an important role for the thermal printhead. Windischmann <sup>3</sup> has reported that the as-deposited SiC films are in compression with absolute values at high as 2×10<sup>10</sup> dyne/cm<sup>2</sup>. The origin of the stress is attributed to hydrogen incorporation, as evidenced by C-H and Si-H bands observed in infrared transmission measurements. Microhardness and scratch adhesion testing are the most commonly used techniques for assessing the mechanical properties of thin surface coatings. Burnett et al <sup>4</sup> have reported schematic representation of coating failure modes in the scratch test. They revealed different film adhesion force results in different failure modes.

In the present study, thermal stress changes for SiC film deposited by PECVD in different gas ratios were investigated. They are correlated with variations of chemical bonding structure for SiC films subtracted out by infrared spectrum (FTIR). Adhesion characteristics of passivation layers, tantalum (Ta) and SiC, played an important role in the printhead thin film

structure. Scratch testing was performed to find out Ta film coating adhesion on the different depositing conditions of SiC films.

## 2. EXPERIMENTAL PROCEDURE

Amorphous SiC films were prepared on silicon wafers of 4-inches diameter. The film were deposited in a capacitively coupled, parallel plate rf glow discharge apparatus. The system (SLR-730, Plasma Therm Inc., FL) has a four-wafers load-lock system, including 125 kHz low frequency (LF) and 13.56 MHz high frequency (HF) rf generators. Deposition experiments were performed at a temperature of 350°C, a power of 100 w, a pressure of 1300 mT and using CH<sub>4</sub>/SiH<sub>4</sub> gas mixtures with helium (He) as a gas carrier. Table I indicates the gas flow ratio of CH<sub>4</sub>/SiH<sub>4</sub>, and other different processing parameters for SiC films using SLR-730 system. Film thickness was measured by the prism coupler. (Meticon Model 2010) Film stress was calculated from wafer curvature measured on the laser-based type at the dual wavelength of 670 nm and 750 nm (FLX-2320, Tencor). Thermal stress changes of SiC films were performed in the nitrogen atmosphere from room temperature to 450°C at a rate of 7.5°C /min, then cooled in the same gas for six thermal cycles. Si-H concentrations were computed from infrared spectra (FTIR, with the Si substrate premeasured and subtracted out).

The adhesion strength of Ta coating (600 nm) on SiC films prepared by different gas ratios from Table 1 at a high frequency (13.56 MHz) were measured by the scratch tester. For distinguishing, scratch testing on Ta over SiC film deposited by a low frequency (125 KHz) was conducted. Parameters for the scratch tester are including of 125 um diamond head, scratch length at 0.3~1.0 cm, loading speed at 1.7 Nt/sec, and the maximum loading force at 10~20 Nt.

## 3. RESULTS AND DISCUSSION

### 3.1 DEPOSITING CHARACTERISTICS OF A-SiC:H FILM

Fig.1 shows that deposition rate of a-SiC:H films decreases by increasing CH<sub>4</sub>/SiH<sub>4</sub> gas ratio. In this experiment, the system is performed at a low operating power region, therefore the ionized threshold for the CH<sub>4</sub> gas is not reached. CH<sub>4</sub> gas molecules would absorb supplied power from PECVD instead of forming reactive radicals. Therefore higher gas ratio caused lower deposition rate for a-SiC:H films.

### 3.2 CHEMICAL BONDING STRUCTURE OF A-SiC:H FILM

Fig.2 shows the typical absorption spectrum for a-SiC:H film measured by FTIR. It indicates those four possible chemical bonds for every absorptive peak on a typical infrared spectrum. For the SiH<sub>n</sub> absorptive peak, wave number showed a little shift ranging from 2000 to 2140 cm<sup>-1</sup> when carbon compositions wt% increased, as shown in Fig.3. This result is caused from different electron negativity between Si and C.<sup>5</sup> In Si-rich region, SiH<sub>n</sub> reveals Si-H or Si-H<sub>2</sub> bonding; but in C-rich region, SiH<sub>n</sub> reveals preferentially Si-H bonding. Si-H<sub>n</sub> bonding concentration can be calculated from the following equation:<sup>6,7</sup>

$$N = A_s \int \frac{\alpha(\nu)}{\nu} d\nu \approx \frac{A_s}{\nu_0} \int \alpha(\nu) d\nu$$

Where A<sub>s</sub> is the inverse absorption cross section of the considered mode, ν<sub>0</sub> is the wave number corresponding to the absorption peak, α(ν) is absorption coefficient. Fig.4 shows Si-H concentration variations with different gas ratios for as deposited and after six thermal cycles.

### 3.3 STRESS ANALYSIS OF A-SiC:H FILM

The total stress in the structure is made up of two terms, i. e., stress due to differential thermal expansion effects and that due to intrinsic stress, such as stress due to argon incorporation, impurity.<sup>2</sup> Thermal stress is resulted from the thermal expansion mismatch effects between a-SiC:H film and Si substrate. Intrinsic stress is due to chemical bonding structure within different compositions of a-SiC:H films. Variations of bonding structure are dependent on different gas ratios. Fig.5 (a) and (b) reveal thermal stress changes of two different gas ratios of a-SiC:H films are performed in the N<sub>2</sub> atmosphere from the room temperature to 450°C. According to the first thermal cycle, stress change is inclined to more compressive stress until 350°C, this result is due to bigger thermal expansion coefficient of a-SiC:H film than Si substrate. Above 350°C, the stress changes toward more tensile according to shrinking of SiC films forming a condense film structure. Therefore the intrinsic stress causes substantially total stress changes toward tensile stress when the temperature is back to the room

temperature. Fig.5 also indicates that total stresses are inclined to saturation according to the small shrinking of SiC films after six thermal cycles. Fig.6 shows that total stresses change toward tensile stresses for every a-SiC:H films deposited from different gas ratios on the Table I. Fig.7 indicates that stress changes after six thermal cycles decrease as CH<sub>4</sub>/SiH<sub>4</sub> gas ratios increased. Thermal stress changes are correlated with chemical bonding structure investigated by FTIR technique. Variations of the Si-H concentration for a-SiC:H films with different gas ratios are shown in Fig.8. Less variation of the Si-H absorption bond causes less thermal stress change.

### 3.4 TA/SIC FILM ADHENSION

Thin-film adhesion is an important problem in thermal ink-jet printhead between the Ta thin film and a-SiC:H films. A qualitative measure of film adhesion can be made with the scratch tester. The adhesive critical load and Ta coating failure modes on a-SiC:H were acquired to examine the film adhesion of these two investigated films. Fig.9 (a) reveals the optical microscopic image for Ta coating over a-SiC:H prepared by a gas ratio of 90 and high frequency of 13.56 MHz. Nevertheless, acoustics emission signal acquired from the scratch tester with scratch length on Ta film is shown in Fig.9 (b). It reveals that critical load is 3.5 Nt. In the present study, critical loads for Ta coating over a-SiC:H films deposited at a high frequency shown in Table I are ranged from 1.6 to 3.5Nt. Failure mode on Ta coating over a-SiC:H film by scratch tester, as shown in Fig.9(a), reveal a crack with regular chipping failure mode, it is a kind of delaminating phenomena. <sup>4</sup> In order to enhance film adhesion of these two investigated films, low frequency power (125 kHz), working pressure (900 mT), and the same gas ratio are used to deposit a-SiC:H film, then the same Ta film coating on it. Fig.10 (a) reveals the optical microscopic image for Ta coating over a-SiC:H prepared by low frequency power, failure mode on Ta is tensile cracking when the coating remains fully adherent. Fig.10 (b) shows acoustics emission signal with the scratch length on Ta film and critical load reaches to 18 Nt as compared with optical microscopic images, shown in Fig.10 (a). Thin-film adhesion was affected by two factors: one is bonding force between Ta and a-SiC:H films, the other is the effective contact area of the interface. Hey et. al. <sup>8</sup> reported that ion bombardment of the a-SiC:H film prepared by low frequency and working pressure can result in more condense film. Fig.11 (a) and (b) show infrared spectra of the high frequency and low frequency a-SiC:H films at the same gas ratio, respectively. Obviously, Si-H and C-H bonds are rarely found out in infrared spectrum of a-SiC:H film by a low frequency. The adhesion depends on the nature of the interfacial region, which depends on the interactions between the depositing Ta thin film and the surface a-SiC : H films. An increased effective contact area in the interfacial region promotes a good adhesion.

### CONCLUSIONS

Higher gas ratio, CH<sub>4</sub>/SiH<sub>4</sub>, caused lower deposition rate for a-SiC:H films. Thermal stress changes of a-SiC:H films were examined in the nitrogen atmosphere from room temperature to 450°C at a rate of 7.5°C /min, then cooled in the same gas for six thermal cycles. Thermal stress changes are correlated with chemical bonding structure investigated by FTIR technique. Less variation of the Si-H absorption bond causes less thermal stress change. a-SiC : H films are sustained by a variety of thermal and mechanical stresses that are detrimental to its integrity. Developing a passivation film for these exacting requirements presented some interesting challenges. Thin-film adhesion is an important problem in thermal ink-jet printhead between the Ta and a-SiC:H passivation films. An increased effective contact area in the interfacial region promotes a good adhesion.

### REFERENCES

1. Eldurkar V. Bhaskar and J. Stephen Aden, "Development of the Thin-film Structure for the ThinkJet Printhead," *Hewlett-Packard Journal* 5, pp.27-37, 1985.
2. L. S. Chang, P. L. Gendler, and J. H. Jou, Thermal, mechanical and chemical effects in the degradation of the plasma-deposited a-SiC:H passivation layer in a multilayer thin-film device, " *Journal of Materials Science* 26, pp.1882-1890, 1991.
3. H. Windischmann, "Intrinsic Stress and Mechanical properties of Hydrogenated Silicon Carbide Produced by Plasma-Enhanced Chemical Vapor Deposition," *J. Vac. Sci. Technol.* Jul/Aug, pp.2459-63, 1991.
4. P. J. Burnett and D. S. Rickerby, "The relationship between hardness and scratch adhesion," *Thin Solid Films* 154, pp. 403-416, 1987.
5. M.A.El Khakani, M. Chaker, A. Jean, S. Boily, H. Pepin and J.C. Kieffer, J. Durand and B. Cros, F. Rousseaux, S. Gujrathi, "Effect of rapid thermal annealing on both the stress and the bonding states of a-SiC:H films," *J.Appl. Phys.* 4, pp. 2834-2840, 1993.
6. C. J. Fang, K. J. Gruntz, L. Ley and M. Cardona, F.J. Demond, G. Muller and S. Kalbitzer, *Journal of Non-Crystalline Solids* 35&36, pp. 255, 1980.

7. Fuminori Fujimoto, Akio Ootuka, Ken-ichiro komaki , Yasushi Iwata, Isao Yamane, Hiroshi Yamashita, Hiroaki Okamoto and Yoshihiro Hamakawa, *Japanese Journal of Applied Physics* 7, pp. 810, 1984.
8. H. P. W. Hey, B. G. Sluijk, D. G. Hemmes, "Ion Bombardment: A determining factor in Plasma CVD, " *Solid State Technology* 4, pp.139, 1990

Table I Processing parameters of a-SiC:H film

CH <sub>4</sub> /SiH <sub>4</sub>	3	6	12	15	30	45	60	90
SiH <sub>4</sub> flow rate (sccm)	450	450	450	450	450	300	225	150
CH <sub>4</sub> flow rate (sccm)	67.5	135	270	337.5	675	675	675	675
He flow rate (sccm)	782.5	715	580	512.5	175	325	400	475
Depositing time (min)	100	100	100	100	100	100	100	150
Film thickness (nm)	1794.6	1014.5	1015	940	1051.5	503.2	633.7	853.3

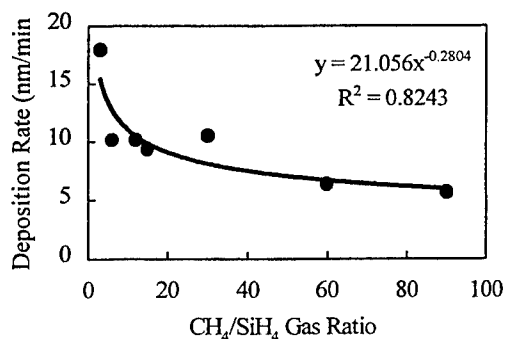


Fig. 1 Deposition rate of a-SiC:H films deposited by different CH<sub>4</sub>/SiH<sub>4</sub>, gas ratio

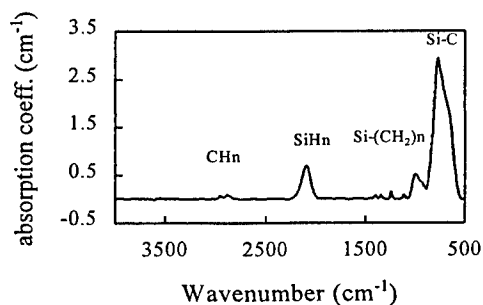


Fig.2 Typical absorption spectrum for a-SiC:H film measured by FTIR

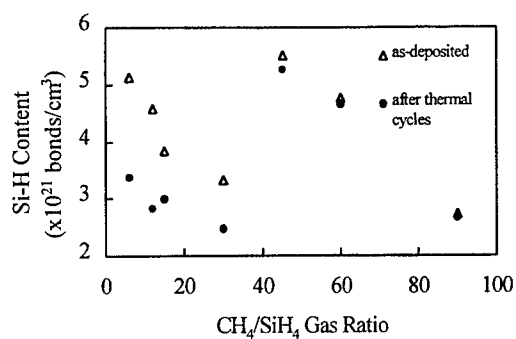


Fig. 3 Si-H<sub>n</sub> wavenumber vs carbon% in a-SiC:H films

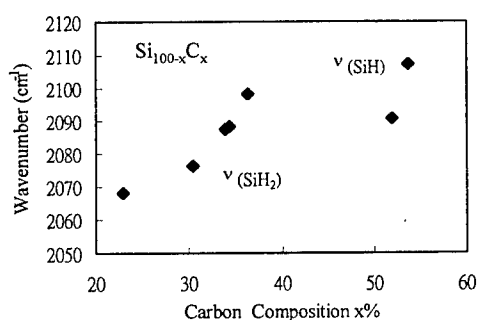


Fig. 4 Si-H concentration vs CH<sub>4</sub>/SiH<sub>4</sub> gas ratio

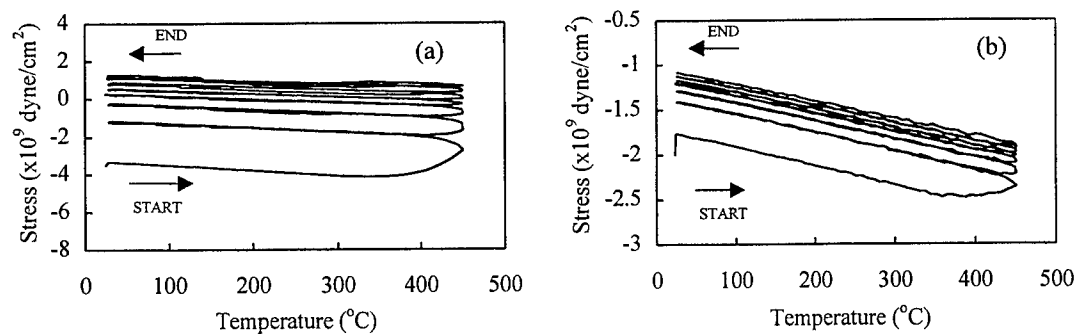


Fig.5 (a) Thermal stress curve for a-SiC:H film (gas ration=6) and (b) Thermal stress curve for a-SiC:H film (gas ration=90)

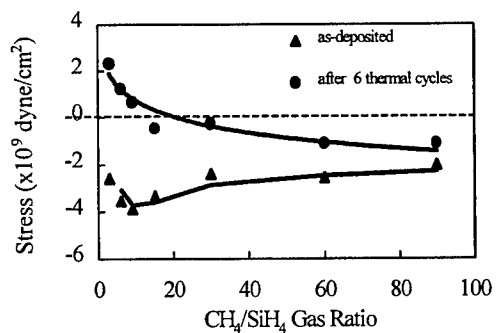


Fig. 6 Film stress before and after 6 thermal cycles for a-SiC:H films prepared by different gas ratios

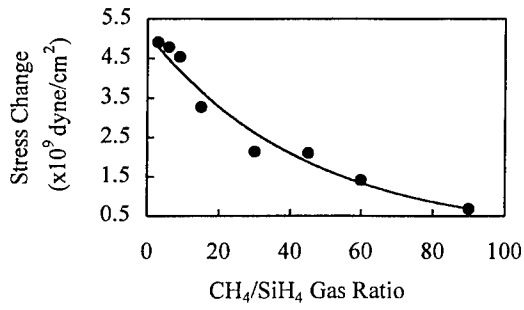


Fig.7 Stress changes after 6 thermal cycles for different gas ratios

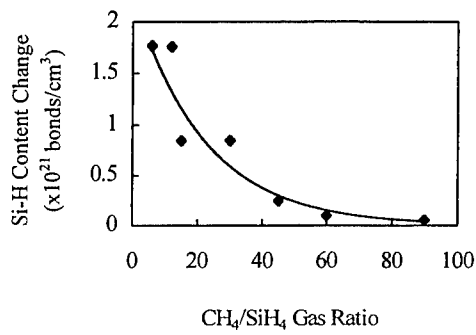


Fig. 8 Si-H concentration change after 6 thermal cycles for different gas ratio a-SiC:H films

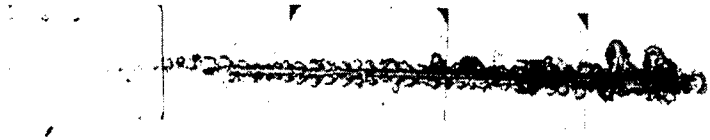


Fig.9 (a) Optical microscopic image for Ta scratch testing on high frequency a-SiC:H film

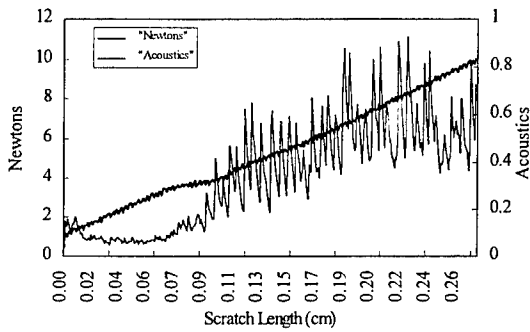


Fig.9 (b) Acoustics emission signal vs scratch length for Fig. 9(a)





Fig. 10 (a) Optical microscopic image for Ta scratch testing on low frequency a-SiC:H film

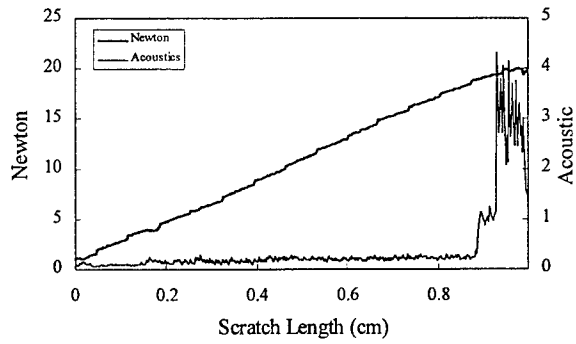


Fig. 10 (b) Acoustics emission signal vs scratch length for Fig. 10 (a)

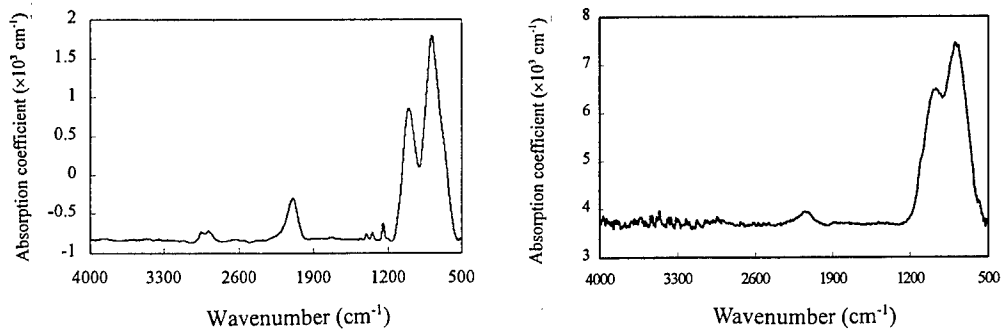


Fig. 11 (a) FTIR spectrum for a-SiC:H film deposited by high frequency, gas ratio=12  
(b) FTIR spectrum for a-SiC:H film deposited by low frequency, gas ratio=12

# A monolithic Thermal Inkjet Printhead Combining Anisotropic Etching and Electro Plating

Chen Yue Cheng, Je-Ping Hu, Yi-Hsuan Lai, Hui-Fang Wang and Chia-Tai Cheng  
Opto-Electronics & System Laboratories, ITRI

## ABSTRACT

The paper proposed a high resolution single-chip monolithic inkjet printhead by combination glowing of nozzle plate on the silicon substrate and anisotropic etching<sup>1</sup>. Ink channels are defined by a sacrificed layer and etched through a mesh network by anisotropic etching. Silicon based channels are strong enough to against the attack of ink solution. Surface planarization is achieved by using PECVD deposited low stress dielectric film on top of the channels to seal mesh cavities. The heater elements are buried in a thin-sandwiched membrane and face down like a back shooter<sup>2</sup>. Ink slot is formed by Etching the substrate from the backside and channel is then connected with Ink slot. Electrical forming nozzle plate will be done later. In this structure, no more another ink channel material is necessary since it is formed on Si substrate. By direct forming nozzle plate on chip, alignment for bonding nozzle plate and chip could be avoided to cost down.

**Keywords:** Monolithic , Anisotropic Etching , Electro plating, Planarization

## 1. INTRODUCTION

Most of the user requirements for printing system is low cost, variable color, high quality, high speed, high resolution, low cost and excellent software support. The Drop-On-Demand (DOD) type printer has several types according to it's ink ejecting mechanism, that is piezoelectric and thermal type. The piezoelectric type ejects an ink droplet by vibration piezoelectric material while thermal type ejects an ink droplet by explosive growth of bubble in the chamber. Thermal inkjet printhead has several advantages compared with other technology such as low cost, high resolution, low noise, and ease of color printing. A core element of the inkjet printer is inkjet printhead, which is a successful product of micromachining technology. It determines the print quality, print speed, and maintenance. Conventional inkjet printhead fabricated by complex process including hybrid technology. The nozzle plate, manufactured separately, should be aligned and attached to the chip one by one with additional bonding process.

In this paper, we propose a new integrated fabrication method for the monolithic thermal inkjet printhead combining silicon micromachining and nozzle plate direct electroplating technologies. The heater elements are buried in a thin-sandwiched membrane and face down like a back shooter. However, to improve the efficiency of fluid supply, individual buried ink channel with non-symmetric throttle or neck is performed. Each ink channel corresponds to a single heat element to avoid any crosstalk issues. Ink slot is formed by Etching the substrate from the backside and channel is then connected with Ink slot. Electrical forming nozzle plate will be done later. In this structure, no more another ink channel material is necessary since it is formed on Si substrate.

## 2. DESIGN AND FABIRICATION

Figure 1 shows a perspective view of the monolithic inkjet head. The realization of buried ink channel array below silicon surface for printhead is achieved by anisotropic etching with KOH solution. Here, a poly silicon (1000Å) sacrificed layer will be deposited on the silicon and define the ink channel area. Considering the efficiency of fluid supply, individual buried ink channel with non-symmetric throttle or neck is designed as shown in Figure 2.

SiC(3500Å) is deposited and defined as a mask, KOH etching solution will etch the silicon to form the V-groove (20um) channel via Poly silicon sacrificed layer. After creating the ink channel, the open pore will be sealed by low stress PECVD dielectric film(12000Å)<sup>3</sup>. This process provides a planar surface for heat element to accommodate on. In addition to the

dimention of pores, the sealing material is also of critical importance. PECVD oxide, PECVD nitride or a combination of both can be used for sealing the structure. Stress and strain problems must be took attention here. Channel to channel space as small as 4um is possible because of the high lateral dimension control of anisotropic etching.

The next step is the creation of conductor and the heater. The heater is just on the diaphragm above the etched channel. After deposition and pattern the Al (6000Å) and TaAl(1000Å), low temperature PECVD dielectrics such as SiN/SiC (5000Å/3000Å) are deposited for passivation. A metal layer (2000Å) for the purpose of both passivation and electroplating seed layer is sputtered on the surface and then patterned using reactive ion etching. Another RIE etching will be applied to create open hole and bonding area for ink droplet and bonding respectively just after metal etching. Electroplating is used to form nozzle plate directly on chip. On the back side of the wafer, the thermal oxide layer (16000Å) is patterned, and a second KOH etching is applied to etch the ink slot. The backside of the wafer should be aligned with the front side. A layer of wax protects front side from the corrosive chemical solution. This slot works as an ink supply chamber and connects to the front side pre-etched ink channel. Figure 3 shows the fabrication process for the monolithic inkjet head.

Prototype roof shooter with 300 dpi resolution and side shooter with 1000dpi have been designed. The realization of buried ink channel and surface planarization is shown in Figure 4. The open pore with 2×2um size is completely sealed. The minimum thickness to seal these pores is about 1.2um. Figure 5(a) and 5(b) show edge type and roof type shooter with heater on ink channel respectively. To edge shooter, wafer must be cut perpendicular to the channel orientation using a dicing machine to expose the nozzles. Figure 6 shows the relating location of electric wire and nozzle. The displacement of heater from nozzle is about 150um(not shown). Figure 7 shows a ink slot view of chip from back side. The front side ink channels connect to this backside ink reservoir. When a drop of ink is fired, the micro channel automatically refills from the reservoir by capillary action.

### 3. CONCLUSION

After the basic process were proven we construct the first sample. A monolithic printhead with high resolution potential has been developed. By direct forming nozzle plate on chip, alignment for bonding nozzle plate and chip could be avoided to cost down. In this structure, no more another ink channel material is necessary since it is formed on Si substrate. By direct forming nozzle plate on chip, alignment for bonding nozzle plate and chip could be avoided to cost down.

It is possible to fabricate both roof shooter and edge shoot by this technique. Right now, printhead with 50 nozzles and system pitch down to 25um (appropriate for 1000 dpi) is evaluating.

### REFERENCE

1. Qingxin Zhang, Litian Liu, Zhijian Li "A new approach to convex corner compensation for anisotropic etching of (100) Si in KOH", The 8<sup>th</sup> International Conference on Solid-State Sensors and Actuators, and Eurosensors IX P.321-P.324.
2. P. Krause, e. Obermeier and W. Wehl "Backshooter-A new smart micromachined single-chip inkjet printhead", The 8<sup>th</sup> International Conference on Solid-State Sensors and Actuators, and Eurosensors IX P.325-P.328.
3. Jingkuang Chen and kensall D. Wise "A High-resolution silicon monolithic nozzle array for inkjet printing",IEEE TRANSACTIONS ON ELECTRON DEVICES. VOL. 44 NO. 9. SEPTEMBER 1997.
4. Jae-Duk Lee, jun-Bo Yoon, Jae-Kwan Kim, Hoon-Ju Chung, choon-Sup Lee, Hi-Deok Lee, Ho-jun Lee, Chong-ki kim, and Chul-Hi Han "A thermal inkjet printhead with a monolithically fabricated nozzle plate and self-aligned ink feed hole", JOURNAL OF MICROELECTROMECHANICAL SYSTEMS. VOL 8. NO.3 SEPTEMBER 1999.

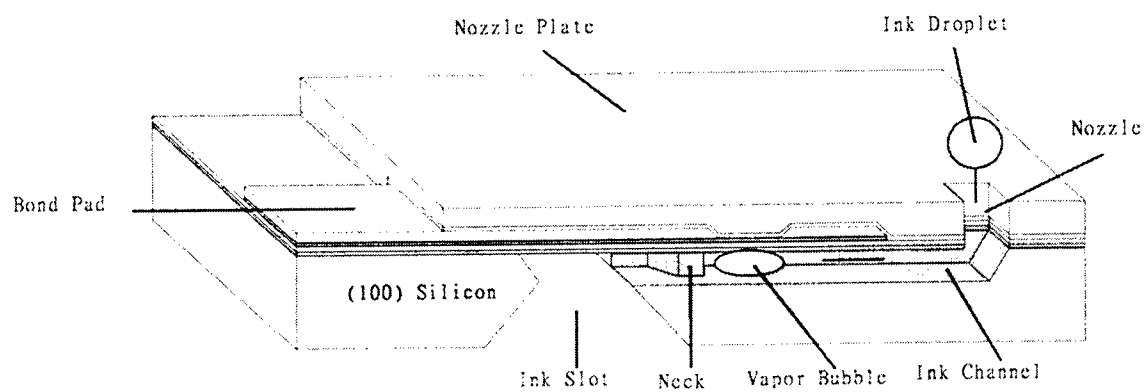


Figure 1. A perspective view of the monolithic inkjet head.

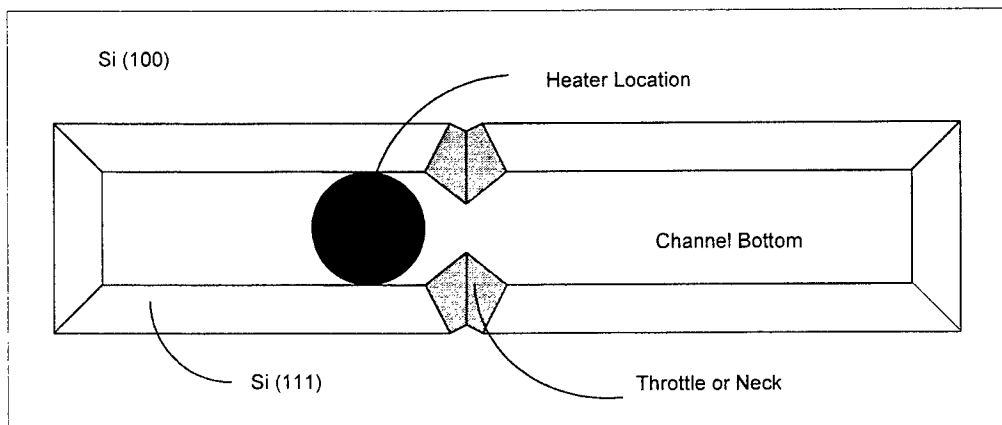
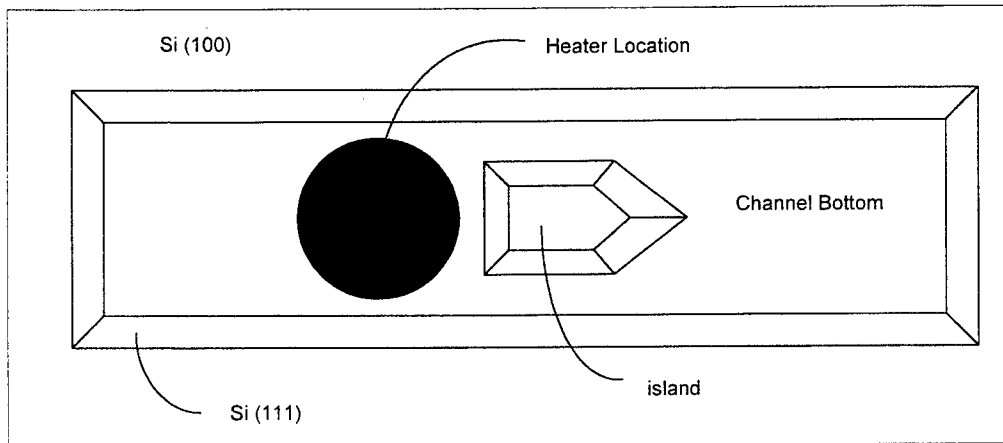


Figure 2. Design of neck or throttle for ink channel.

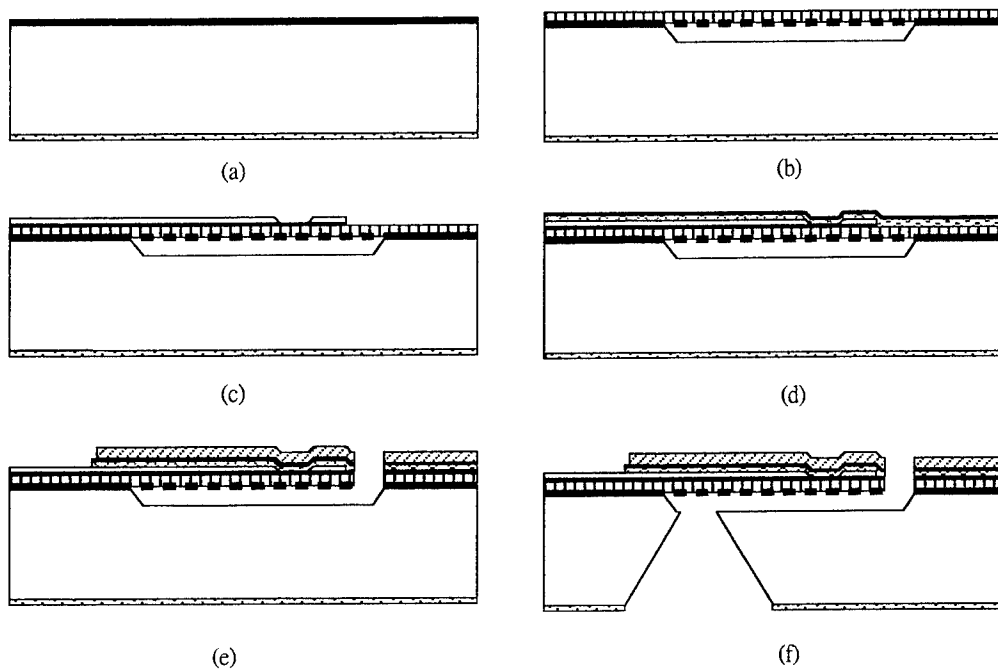


Figure 3. Process flow for the monolithic inkjet head. (a) Ink channel defined and protect by dielectric film. (b) Ink channel anisotropic etched and open pole sealed. (c) Conductor and heater creating. (d) Passivation deposition. (e) Pad area, nozzle hole define and nozzle plate electroplating. (f) Back side ink slot etching.

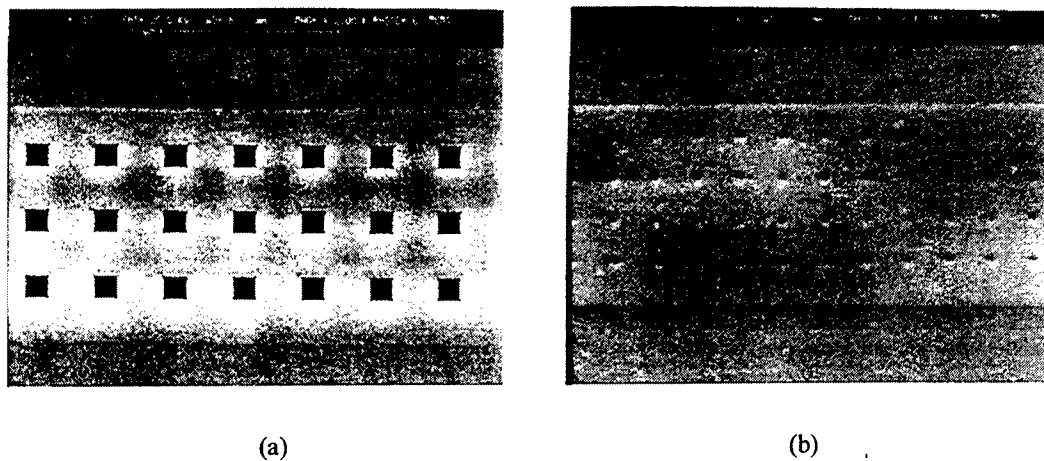
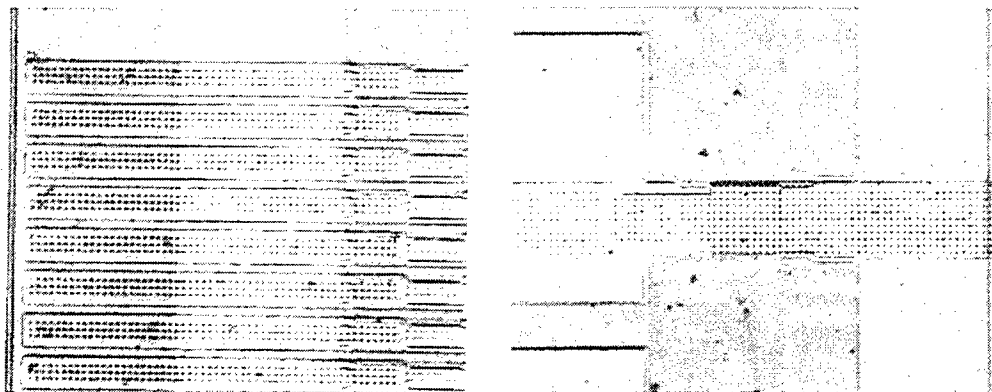


Figure 4. (a)  $2\mu\text{m} \times 2\mu\text{m}$  open pores. (b) Pores is sealed by PECVD deposition.



(a)

(b)

Figure 5. (a) 1000dpi edge side shooter. (b) 300 dpi roof shooter.

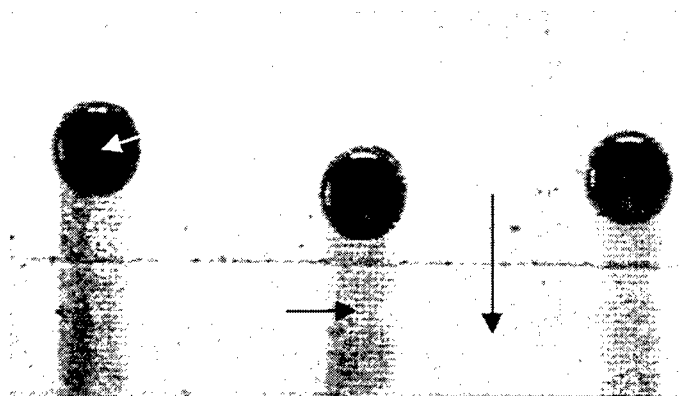


Figure 6. The relating location of nozzle and electric wire.



Figure 7. A ink slot view of chip from back side



# Structured-Light Based on Shaped Depth-image Capturing System

Chen Deyun<sup>a</sup>, Tan Guangyu<sup>a</sup>, Yu Xiaoyang<sup>b</sup>, Meng Qingxin<sup>b</sup>

<sup>a</sup>Harbin University of Science and Technology, Harbin 150080, China

<sup>b</sup>Harbin Engineering University, Harbin 150080, China

## ABSTRACT

At first, this paper introduces the structure of the depth-image capturing system using the single-stripe pattern. . At second, its operating principle is introduced, its mathematical model is established and the calibrating method for it is put forward. At last, its prototype is produced and calibrated.

**Keywords:** Structured-light, Depth-image

## 1. INTRODUCTION

It is not unique to recover the 3D shape of objects from their usual intensity images because information of depth is lost. But the depth-images have 3D information of object surfaces, so they are used not only to measure 3D shape of object surfaces but also to provide a new approach to pattern recognition.

This paper puts forward a kind of depth-image capturing system based on structured-light. As known, its distortion can be transformed into the height change in the direction of the stripe if a single-stripe light is emitted and observed sideways. So a generator emitting a single-stripe light and a camera can make up of a depth-image capturing system. It can obtain the depth-image of the detected object section. If the object is shafted by definite step along a fixed beeline direction( $x$ ), the depth-image of full object surface is achieved when the system repeats the above said operation<sup>1</sup>.

## 2. THE BUILDUP OF OUR SYSTEM

The buildup of the depth-image capturing system based on structured-light this paper is shown in Fig.1. It includes semiconductor laser and its optical system, CCD camera, image collecting card, computer and its software.

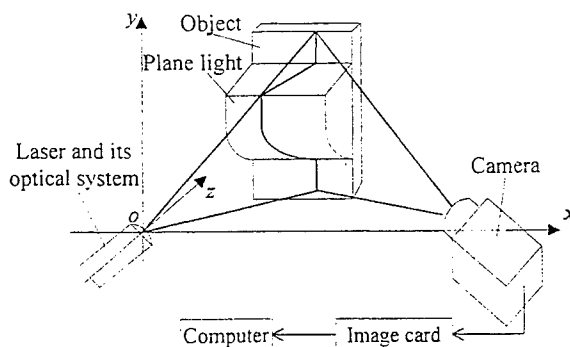


Fig.1 buildup of the depth-image capturing system

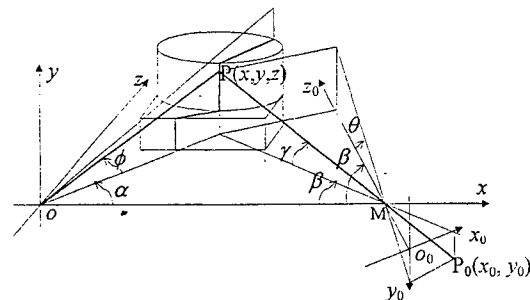


Fig.2 Principle of the system

## 3. MATHEMATICAL OF OUR SYSTEM

The operating principle of our system is shown in Fig.2. The center of the camera len is the point( $M$ ). The light emitted from the laser is becomes a sector plane light through the optical system, which is vertical to the plane( $xoz$ ). Its vertex is located at the origin( $o$ ) of the world coordinate system( $xyz$ ), and  $oM=B$ . The plane light forms a strip on the measured object. Any object point  $P(x, y, z)$  on the strip produces its image point  $P_0(x, y)$  on the image plane( $x_0o_0z_0$ ) in the camera coordinate system( $x_0y_0z_0$ ). The plane( $x_0o_0z_0$ ) coincides with the plane( $xoz$ ). The ray( $oP$ ) is at an angle of  $\phi$  to the plane( $xoz$ ), its projection on the plane( $xoz$ ) is at an angle of  $\alpha$  to the axis( $ox$ ). In this paper,  $\alpha$  is made to equal  $90^\circ$ . The optical axis( $o_0x_0$ ) of

the camera is at an angle of  $\beta_0$  to  $Mo$ . The focus of the camera is equal to  $o_0M$ . The ray(MP) is at an angle of  $\gamma$  to the plane( $xoz$ ), the projection of  $\gamma$  onto the plane( $y_0o_0z_0$ ) is  $\theta$ . The projections of the ray(MP) onto the plane( $xoz$ ) is at an angle of  $\beta$  to  $Mo$ . The visual angle of the camera is  $2\beta_1$ . The pixel sequence numbers of camera along the horizontal direction( $x_0$ ) and the vertical direction( $y_0$ ) are respectively  $n$  and  $m$ , where  $n$  is from  $-N$  to  $N$  and  $m$  from  $-M$  to  $M$ .

According to Fig.2, we have from the camera geometry<sup>2</sup>

$$z = \frac{B}{\frac{ctg\beta_0 + \left(\frac{tg\beta_1}{N}\right)n}{1 - \left(\frac{ctg\beta_0 tg\beta_1}{N}\right)n}} = B \frac{1 - \left(\frac{ctg\beta_0 tg\beta_1}{N}\right)n}{ctg\beta_0 + \left(\frac{tg\beta_1}{N}\right)n} \quad (1)$$

$$x = z \cdot ctg\alpha = 0 \quad (2)$$

$$y = \frac{z \left(\frac{tg\beta_1}{M}\right)m}{\sin\beta_0 - \cos\beta_0 \left(\frac{tg\beta_1}{N}\right)n} \quad (3)$$

where  $n$  and  $m$  are obtained from the image; the parameters of  $\beta_0$ ,  $\beta_1$ ,  $B$  are constants to be calibrated.

#### 4. METHOD OF OUR SYSTEM CALIBRATING

##### 4.1. Calibrating Model of Our System

The calibrating principle of our system is shown in Fig.3. Fig.3 is the projection of Fig.1(Fig.2) on the plane of  $xoz$ . In Fig.3, The reference plane is parallel to the plane of  $xoy$  and the distance between them is equal to  $A$ ; the distances from object points of  $Z_1$ ,  $Z_2$  and  $Z_3$  to the reference plane are  $z_1$ ,  $z_2$  and  $z_3$ , their image points are separately on the right side column, the middle column and the left side column of the camera image plane.

According to the geometrical relation in Fig.3, we have

$$z_1 + A = B \cdot \tan(\beta_0 - \beta_1) \quad (4)$$

$$z_2 + A = B \cdot \tan\beta_0 \quad (5)$$

$$z_3 + A = B \cdot \tan(\beta_0 + \beta_1) \quad (6)$$

From formula (4), (5) and (6), the following formula is deduced.

$$B^2 = \frac{(z_1 + z_3 + 2A)(z_2 + A)^2 - 2(z_1 + A)(z_2 + A)(z_3 + A)}{z_1 + z_3 - 2z_2} \quad (7)$$

It is clear that  $B$  can be obtained if a reading of  $A$  is given. Formula (7) is the calibrating model of  $B$ .

From formula (4), (5) and (6), the following formulas are deduced again.

$$\beta_0 = \arctan \frac{z_2 + A}{B} \quad (8)$$

$$\beta_1 = -\arctan \frac{z_1 + A}{B} + \beta_0 \quad (9)$$

or

$$\beta_1 = \arctan \frac{z_3 + A}{B} - \beta_0 \quad (10)$$

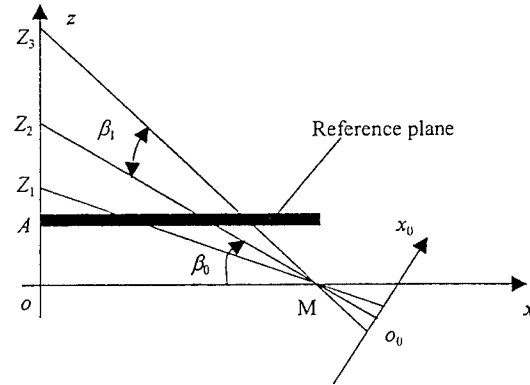


Fig.3 Projection of our system on the plane of  $xoz$

Formula (8) and (9) or (10) is the calibrating model of  $\beta_0$  and  $\beta_1$ .

From formula (3), we have

$$y = \frac{m}{M} \cdot \tan \beta_2 (B \cdot \cos \beta_0 + z \sin \beta_0)$$

When column number of  $n$  is changeless, we have

$$\cot \beta_2 = \frac{m_1 - m_2}{y_1 - y_2} \cdot \frac{B \cdot \cos \beta_0 + z \cdot \sin \beta_0}{M} \quad (11)$$

It is clear that  $\beta_2$  can be obtained if readings of  $(m_1 - m_2)$  and  $(y_1 - y_2)$  are given. Formula (11) is the calibrating model of  $\beta_2$ .

#### 4.2. Calibrating Process of Our System

The calibrating equipment includes a guide with a scale, placed along the lines of  $oz$ , and a standard plane parallel to the reference plane, placed on the moving part of the guide. At first,  $z_1$ ,  $z_2$  and  $z_3$  are calibrated across experiment. And then,  $B$ ,  $\beta_0$ ,  $\beta_1$  and  $A$  are calibrated. At last,  $\beta_2$  is calibrated. The concrete process is as the following:

- 1) The standard plane is shifted until the single-stripe image appears on the right slide column of image plane. Here the standard depth  $z_1$  is 111mm according to the scale of the guide.
- 2) The standard plane is shifted until the single-stripe image appears on the middle slide column of image plane. Here the standard depth  $z_2$  is 320mm according to the scale of the guide.
- 3) The standard plane is shifted until the single-stripe image appears on the left slide column of image plane. Here the standard depth  $z_3$  is 910mm according to the scale of the guide.
- 4) The designed value of  $A$  ( $A=95\text{mm}$ ),  $z_1$ ,  $z_2$  and  $z_3$  are substituted into formula (7), (8), (9) or (10). Then the values of  $B$ ,  $\beta_0$  and  $\beta_1$  are calculated.
- 5) The standard plane is shifted until the single-stripe image appears on some column of image plane. Here the standard depth  $z_i$  is obtained from the scale of the guide and the reading of  $n_i$  is given by the camera image plane.
- 6) Many groups of  $(z_i, n_i)$  are obtained when repeating process 5).
- 7) According to  $n_i$ , the depth calculated value of  $z'_i$  is obtained from formula (1).

- 8) Calculate least square error of  $\sum_{i=1}^l (z_i - z'_i)^2$ .

- 9) Correct the value of  $A$  to make the least square error reduce.

- 10) Repeat process 4)~9) until the least square error becomes least. Here the value of  $A$  is best and used as calibrated value.

The calibration result is as the following:  $B=311.916\text{mm}$ ,  $\beta_0=52.7372^\circ$ ,  $\beta_1=19.9293^\circ$ ,  $A=90\text{mm}$ .

Row numbers of two images of the same column of  $m_1$  and  $m_2$  can directly be obtained from the image. Then the value of  $\beta_2$  can be calibrated according to formula (11), if the height difference of  $(y_1 - y_2)$  is known, where the object point height corresponding to image point row number  $m_1$  is expressed as  $y_1$  and  $m_2$  expressed as  $y_2$ . The concrete process is as the following:

- 1) The standard measuring block is placed in the view field of the system, perpendicular to the plane of  $xoz$  and parallel to the reference plane, shown as in Fig.4.
- 2) The upper surface height of the standard measuring block is expressed as  $y_1$ , the lower surface height expressed as  $y_2$ . Then the standard value of the measuring block (100mm) is used as the height difference  $(y_1 - y_2)$  in formula (11).
- 3) The intensity image of the block is shown in Fig.5. The row numbers of  $m_1$  and  $m_2$  are made sure by image processing. The two-valued fringe image of the block is shown in Fig.6. From its grey data,  $m_1$  and  $m_2$  can directly be obtained.
- 4) The calibrated value of  $\beta_2$  can be calculated when substituting  $(y_1 - y_2)$  and  $(m_1 - m_2)$  into formula (11).

The calibration result is that  $\beta_2=15.0703^\circ$ .

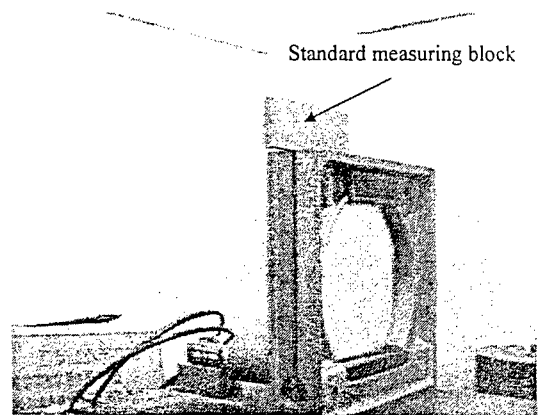


Fig.4 Placement of the standard measuring block

## 5. EXPERIMENT RESULTS

As shown in Fig.7, the bright stripe is the section outline of a plaster head portrait figure, and its depth image is shown in Fig.8.

The depth image experiment of a water heater model is shown in Fig.9. The stripe image of the model is as shown in a, and the depth image of the model is as shown in b.

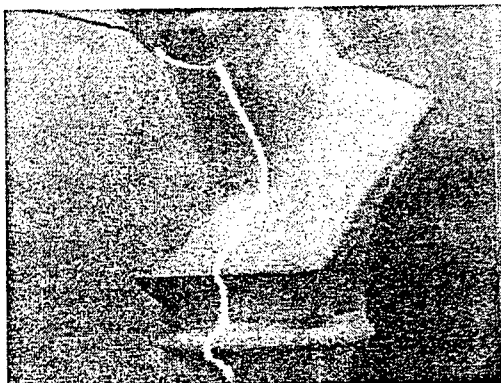


Fig.7 Stripe image of the plaster head portrait



Fig.5 Grey image of block

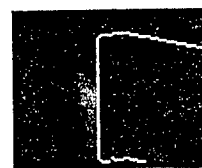


Fig.6 Two-valued fringe image of block

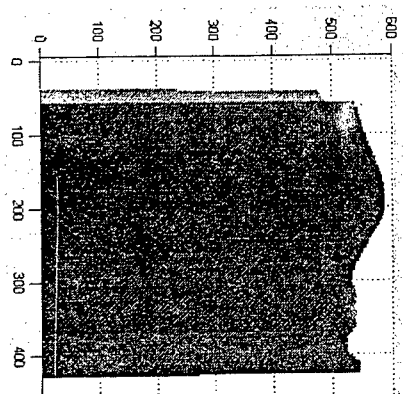
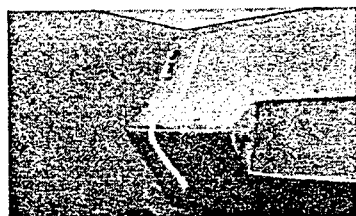
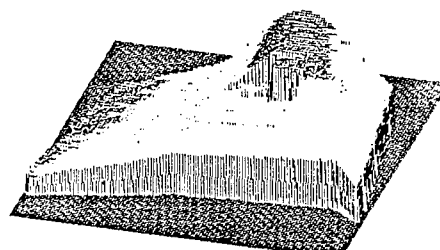


Fig.8 Depth image of section of the plaster head portrait



a. Stripe image of the water header model



b. Depth image of the water header model

Fig.9 Depth image experiment result of the water header model

## 6. CONCLUSION

This paper introduces the depth image capturing system based on structured-light and proposes the calibration method for our system. The experiment results show that our system prototype has the measuring scale of  $300\text{mm}(y) \times 300\text{mm}(z)$ , the resolving power of  $0.5\text{mm}(y) \times 0.5\text{mm}(z)$  and the precision of  $1.5\text{mm}(y) \times 1.5\text{mm}(z)$ .

## ACKNOWLEDGMENTS

This work is supported by Natural Science Foundation of China(NSFC) and Natural Science Foundation of Heilongjiang Province(NSFH).

## REFERENCES

1. Fu Jingxin, **Robotics**, pp.194-200, China Press of Science and Technology, Beijing, 1989
2. X. Yu, J. Zhang, L. Wu, X. Qiang, "Laser Scanning Device Used In a Space-Encoding Rangefinder," pp.348-351,



## Author Index

- Chang, Horng, 55  
 Chang, I-Cheng, 21  
 Chang, Kuo-Shu, 192  
 Chen, Bor-Tow, 21  
 Chen, Chien-Chung, 87  
 Chen, Chin-Tai, 64, 214, 222  
 Chen, Chun-Yen, 136, 159  
 Chen, Deyun, 253  
 Chen, Tsuhan, 8  
 Cheng, Chen-Yue, 239, 246  
 Cheng, Chia-Tai, 246  
 Cheng, Fang-Hsuan, 167  
 Cherng, Ya-Tung, 55  
 Chiang, Hwang-Cheng, 112  
 Chiu, Ching-Long, 72  
 Cho, Gang Seok, 200  
 Chuang, Kai-Wai, 96, 104  
 Hiruma, Teruo, 2  
 Hong, Hong-Ming, 55  
 Hsieh, Kun-Jiang, 21  
 Hsu, Wei-Feng, 96, 104  
 Hsu, Yun-Chiang, 96, 104  
 Hsueh, Wen-Jean, 14, 21, 78  
 Hu, Hung-Lien, 230  
 Hu, Je-Ping, 246  
 Huang, Ku-Chung, 78  
 Huang, Shin-Wen, 208  
 Jan, Meei-Ling, 208  
 Jeong, Hun, 200  
 Kim, Chung Hwa, 200  
 Kim, Han Chil, 200  
 Kim, Hong Bin, 200  
 Kim, Hyung-Bum, 180  
 Koh, Sung Shick, 200  
 Kuo, Chung Jung, 29  
 Lai, Chien-Chang, 214  
 Lai, Yi-Hsuan, 230, 246  
 Lan, Yuan-Liang, 72  
 Lee, Bore-Kuen, 136  
 Lee, Hsien-Che, 122  
 Lee, Jyh-jiun, 148  
 Lee, Koo Young, 200  
 Lee, Kuen, 78  
 Lee, Ming-Ling, 230  
 Lee, Yih-Shing, 239  
 Li, Yi-Wei, 192  
 Lian, Zhi-Ru, 230  
 Liang, Hsing-Ching, 208  
 Liaw, Yi-Ching, 159  
 Lim, Chun-Hwan, 180  
 Lin, Hsien-Chang, 21  
 Lin, Tsang-Gang, 29  
 Liu, Hong-Chih, 208  
 Lu, Fu-Fa, 55  
 Meng, Qingxin, 253  
 Nguyen, Minh-Chinh, 41  
 Ni, Catherine W., 48  
 Okuda, Masahiro, 8  
 Park, Jong-An, 180  
 Park, Seungjin, 180  
 Pei, Cheng-Chih, 208  
 Shen, Day-Fann, 192  
 Shen, Feng-Chi, 136  
 Shyu, Chuen-Shing, 208  
 Tan, Guangyu, 253  
 Tang, Jiy-Shan, 208  
 Tsai, Chao-Hsu, 78  
 Tyler, Christopher W., 87  
 Wang, Chiehwen, 72, 230  
 Wang, Hui-Fang, 246  
 Wang, Sheng-Jyh, 112  
 Wen, Chao-hua, 148  
 Wu, Yi-Yung, 72, 239  
 Wu, Dong Sing, 239  
 Yan, Loon-Shan, 192  
 Yang, Chih-Yuan, 167  
 Yeh, Ching-Kai, 208  
 Yeh, Ruey-Nan, 55  
 Yeh, Yeou-Min, 112  
 Yu, Xiaoyang, 253